

Ministry of Education and Science of Ukraine

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic
Institute"

Glushkov Institute of Cybernetics, National Academy of Sciences of Ukraine

Polytechnic Institute of Tomar (Portugal)

Smart Cities Research Center (Portugal)

MODERN ASPECTS OF SOFTWARE DEVELOPMENT

**Proceedings of
VII Scientific and Practical Virtual Conference of
Software Development Specialists**

June, 1, 2020

Kyiv, Ukraine

Modern Aspects of Software Development: Proceedings of VII Scientific and Practical Virtual Conference of Software Development Specialists, June, 1 2020 p. – Kyiv: Igor Sikorsky KPI, 2020. – 83 p.

The Proceedings of VII Scientific and Practical Virtual Conference of Software Development Specialists “Modern Aspects of Software Development” consists of the results of scientific researches and practical developments in the following directions: software engineering, computer ecological-economic monitoring, computerized design

Text is printed in the authors' edition.

ISBN

© Authors' texts, 2020

CONTENTS

<i>Bochok V.</i> Construction of data collection and analytics systems for IT vacancies.....	5	<i>Shaldenko O., Hulak O.</i> Analysis of monitoring methods and identification of illegal outdoor advertising.....	62
<i>Gaidarzhy V., Huchenko M.</i> Web-system for centralized management of distribution educational literature.....	9	<i>Sydorenko IU., Kryvda O.V., Horodetskyi M.</i> Processing of large arrays of data by Gaus interpolation methods.....	66
<i>Gruts YU., Holotiuk P.</i> Implementation of procedures for building wire-frame stereoscopic images with multi-angle observation.....	13	<i>Tytenko S., Omelchenko P.</i> Concept maps in the individual learning environment	71
<i>Husyeva I.</i> Mobile sensors in intelligent transportation systems.....	17	<i>Tytenko S., Talakh O., Zinkevych B.</i> Automated construction of test tasks based on decomposition of educational text.....	75
<i>Karaieva N., Varava I.</i> Methodology design of monitoring system of indicators of the human development level of Ukraine's regions.....	22	<i>Yerokhina A.</i> Comparison of binary methods of multi-label classification.....	78
<i>Kublii L., Horodetskyi M.</i> The use of images for steganographic information protection.....	27		
<i>Kublii L., Mykhal'ko V.</i> Software tools for intellectual and statistical processing of hydroacoustic signals.....	31		
<i>Kuzminykh V., Kotsubanov G.</i> The decision-support platform for international scientific and technical cooperation.....	35		
<i>Kuzminykh V., Polishchuk K.</i> Interpreter of advanced database management commands.....	38		
<i>Kuzminykh V., Taranenko R.</i> Expert system for assessing the quality of information.....	41		
<i>Kuzminykh V. Tymoshenko M.</i> Automation of assessment and forecasting of development of higher educational institutions...	45		
<i>Mykhailova I., Tyschenko A.</i> Web registry of electronic information resources.....	48		
<i>Segeda I.</i> Cloud data storage services.....	52		
<i>Shushura O., Havrylko Y.</i> Conceptual modeling of fuzzy control.....	58		

CONSTRUCTION OF DATA COLLECTION AND ANALYTICS SYSTEMS FOR IT VACANCIES

Modern information technology solves a very wide range of problems, and some of them did not exist even in the last decade. For example, in this decade, tasks such as distributed calculations of large amounts of data, building systems that can handle heavy loads, blockchain, and machine learning come to the fore. New areas continue to emerge and bring new specialized technologies and approaches. But what unites them — they need software implementation that requires knowledge of certain algorithms, principles, and tools. Due to the limited number of engineers and the areas of development that are often strongly related, it requires a specialist to have extensive expertise in many areas.

Business development leads to the evolution of problems that need to be solved. As a result, certain technologies and tools are becoming obsolete and less relevant. In their place come others. This transition takes time, so it is extremely important to be able to analyze trends and make timely decisions in order to always be competitive in the labor market.

Currently, there are services that provide information on the relevance of a limited number of technologies, mostly programming languages, as well as job search sites that provide some surface analysis of certain types of vacancies. The problem of the first type of services, such as the TIOBE index, is the focus on one specific metric and a limited number of technologies [1]. For example, TIOBE focuses on the number of search queries on the Internet. It is quite difficult for ordinary engineers to draw certain conclusions from such ratings. For example, C is ranked first with a rating of 17.7%, and Python is ranked third with a rating of 9.12%, however, there are far fewer vacancies related to the former.

The drawback of analytics that provides job search sites is that, for the most part, such resources work with a wide range of professions, each of which requires certain skills. For such sites, it makes no sense to distinguish certain technologies, such as programming languages, from other text, because for managers or designers they simply do not matter. And vacancies on such resources are filtered using full-text search. Also, such sites can use not the vacancies themselves as a source of information, but surveys of engineers, which in some cases are not representative. A sample of such surveys per year can only recruit 5,000 respondents. We should also take into account that engineers have no motivation to conduct surveys or provide reliable information.

To solve this problem, the subject area was analyzed and it was

decided to build a system based on vacancies as the most representative source. The problem was that the sources of vacancies, although mostly open, were unstructured and contained different formats. Also, a vacancy is a textual description that contains a large amount of information in a random order, which is extremely difficult for a computer to process. In addition, all data is in different sources and must be stored in its own database in order to perform complex operations on them relatively quickly.

The system is complex and consists of several services that work together. The advantage of this approach is the flexibility and ability to change some modules without changing other services, except in situations where the service interface is changing. It also gives an ability to scale the system to respond to heavy loads. For this aim was also implemented OAuth2 authentication, which will reduce a number of queries to the database.

We should also cover a problem of communication for multiple parts of application and easy way to manage the infrastructure to deploy software.

During the modeling process was decided to develop a system, consisting of 3 services and code, which is made in a separate module, but possible to because a separate service in the future to make the code a little bit slower but more flexible.

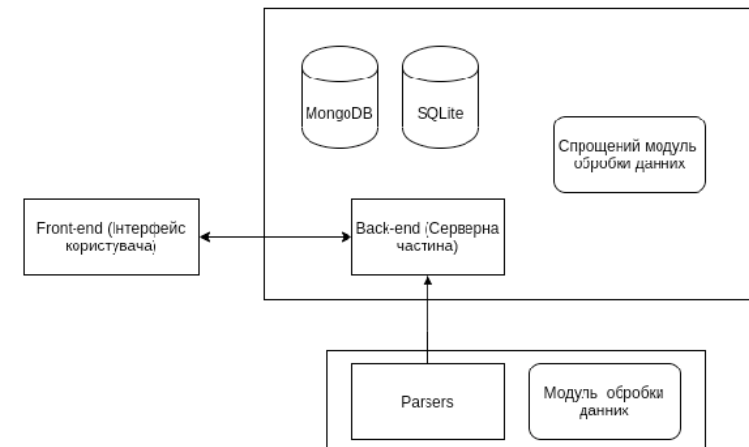


Figure 1 — The structure of interaction of system services

Parsers run every day and process all vacancies currently available. Usually, resources are provided only by vacancies that were active no more than a month ago. Each parser must provide the following data:

1. `id_external` — a string containing a unique identifier from the system from which the vacancy was taken. Required field.

2. `data_source` — string identifies the unique name of the service-source of vacancies. Required field.

3. `url` — path to the vacancy.

4. `title_raw` — vacancy title in raw form.

5. `publication_date` — vacancy publication date (often unavailable field).

6. `date_parsed` — the date when the vacancy was parsed.

7. `required_technologies_raw` — the text of the vacancy (or a certain part of it, which contains information about the technologies that the worker needs to know. It is often very problematic to select the necessary part of the text, so the entire description of the vacancy is saved for analysis).

The next step will be to correct the so-called data extractors. The next step is data transformation covered by data extractors. Their behavior is controlled by the configuration, provided in JSON format. Extractors are used to extract technologies and vacancy levels. They are built to solve most of the collisions that are possible to paper in the text. The full list of fields added by extractors are listed below:

1. `required_technologies` – extracted from vacancy text.

2. `vacancy_level` – extracted from vacancy title.

3. `core_technologies` – extracted from vacancy title.

All collected vacancies are stored in the `unverified_vacancies` collection in MongoDB. This solution allows for some analysis or aggregation of the data. For example, you can add functionality that allows you to view vacancies and correct information manually. It will also allow you to adjust system configurations. The data extractor requires rules encoded in JSON that it can obtain from the server part.

To create a server that will be able to communicate and control all the others was decided to build the unified RESTful API. One of the priorities was to achieve a flexible system with minimum time spent. That is why was used a framework with a great variety of ready-to-use extensions. The Flask was used because of the minimum functionality out of the box with the ability to the extent. The application architecture uses Flask-Restplus [2, 3].

To create a server that will be able to communicate and control all the others was decided to build the unified RESTful API. One of the priorities was to achieve a flexible system with minimum time spent. That is why was used a framework with a great variety of ready-to-use extensions. The Flask was used because of the minimum functionality out of the box with the ability to the extent. The application architecture uses Flask-

Restplus [2, 3] which provides an advanced way to document the app with Swagger and validate user input with Marshmallow. To control the versions of the database was decided to use a Flask-Migrate plugin. SQLAlchemy and PyMongo integrations were added to communicate with the databases.

MongoDB is NoSQL database, which requires a denormalization process performing on data. It's really fast for reading/writing operations and able to store a schema-free data. On the other hand, the engineer needs to control data consistency. For example, if the administrator changes the name of the technology, you need to change this name in all vacancies that are linked to this technology.

There was decided to use both SQL and NoSQL databases to get the best from them. SQLite stores all relational data except the vacancies to control the data consistency and MongoDB is performing complex data filtering and aggregating steps. It's also able to handle a great amount of data and scale horizontally by just adding additional nodes with master-slave replication. It's also good for logging or metrics that can be collected in the future.

The interface was implemented using the React framework. It aims to build interfaces on the principle of single responsibility. The application follows the principles of SPA (single page applications), which means that the application does not have to be rebooted to display new content. Even if it's a new page. However, the link navigation history must be stored by the browser. This approach significantly speeds up the site, because you do not need to download from the HTML server, styles, and libraries for each new request, but only the data that needs to be displayed. It also makes the application much more interactive.

The last step is to configure the services to work one with another. With the docker-compose were described as infrastructure as code, which means that all configurations that are required to be done on app starting are now formalized in the YAML configuration file. All is needed to launch the whole app is to perform a simple terminal command. An additional advantage is that all services will be wrapped in isolated containers and the only way to interact with the app is sending requests to specified ports.

References:

1. TIOBE Programming Language Popularity Index [Electronic resource] - Resource access mode: <https://www.tiobe.com/tiobe-index/>.

2. Flask Framework Documentation [Electronic resource] - Resource access mode: <https://flask.palletsprojects.com/en/1.1.x/>.

3. Fielding R. Architectural Styles and the Design of Network-based Software Architectures: dis. Dr. comp. Science / Fielding Roy.

Gaidarzhly V., Huchenko M.

WEB-SYSTEM FOR CENTRALIZED MANAGEMENT OF DISTRIBUTION EDUCATIONAL LITERATURE

The amount of information and the speed which this data needs to be updated is constantly growing to keep it up to date. Technological problems of the educational process require the availability of an information system for storage and distribution of educational literature, which should contribute to the improvement of the educational process.

Most of the existing tools offer solutions [1] that require a lot of time and resources to implement, which not every educational institution can afford. One of the goals was to create a system that automates the process of sharing educational literature, making it easy to use and easy to implement. Practice shows that some educational institutions use online platforms for publishing training courses or simply use cloud storages to solve some of the problems. In some cases, the use of cloud storage is implicit, while use e-mails.

First of all, the developed system relieves the teacher of responsibility in the management of the distribution of educational materials. The system undertakes to send students books, abstracts, laboratory works and other educational materials on the exact day and time. The system improves the quality and speed of students' mastery of educational materials, compared to traditional approaches.

Paying attention to the fact that the institution of higher education can use non-standard solutions for data storage, the architecture of the developed system takes this into account.

The main focus is on improving the use of educational materials by students with a minimum cost of implementing the system. The main focus is on students studying on a full-time basis and need regular interaction with the teacher.

The following problems need to be addressed when developing an appropriate information system:

- lack of opportunity to plan in advance the date of distribution of educational materials,
- lack of automated tools for the teacher to form an array of educational literature (filling e-mails with identifiers, links to the topic of the curriculum),
- lack of control over the process of mastering educational literature,
- determining the direction of educational materials received by the student over time,
- lack of organizational principles for the preservation of educational

materials.

- It is important to consider a number of issues in more detail.

The cost of implementation

The developed information system allows it to be easily implemented due to the fact that the student does not have to use the system, it is enough to have mail, which will receive all the necessary information with content partially generated by the system and partly by the teacher.

Lack of a unified approach to the design of e-mails

Over time, e-mails receive a large number of letters. Simply, when it comes time to review materials that will be sent over several years, it becomes very problematic. The reason for this is the lack of systematization.

The developed information system uses a single structure for e-mails. This structure adds the following information to the letter:

1. Subject
2. Name of the material
3. Title of the section related to study material
4. The name of each topic related to study material
5. Sender information,
6. Additional information provided by the sender.

All created emails saved into the system can be reused, just need to pick one from the created (figure 1).

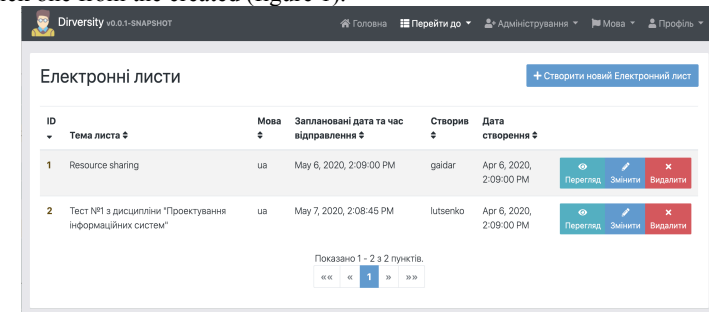


Figure 1. Created emails page

Thanks to the chosen structure, each student can arrange the letters as he wishes and find the materials he needs. This is extremely important because easy access to educational information is one of the highest priorities.

Architecture

The choice of software architecture is always one of the most

important issues when creating an information system. Experience shows that a well-designed architecture can save a lot of time and resources, while a bad one can be fatal even for a project with a fascinating idea.

Software architecture refers to making a set of strategic technical decisions related to a software product, documenting them and ensuring their implementation. Architecture design has become one of the most important stages in building software, especially in the field of development of high-load systems.

It was decided to use the “Monolith First” approach of building the architecture of the web-system for centralized management of distribution educational literature. It allows to begin its development fairly quickly and do not need to spend additional resources to support individual services.

Microservice architecture is important for flexible development, but their use has a significant advantage only for more complex systems. Management of a set of services will significantly slow down the development, which indicates the advantage of using a monolith for simpler systems [3]. This leads to a strong argument of the strategy that it is worth starting the development using a monolithic architectural style, even if the project will later benefit from the architecture of microservices. In this case, the monolith should be designed carefully, paying attention to the modularity of the software, both at the boundaries of the API and data storage. While maintaining the modularity within the monolith, the transition to a microservice architecture is a relatively simple operation.

Therefore, this approach, known as “Monolith First”, will allow you to use all the benefits of a monolithic architecture at the beginning of the system. The next advantage is when there is a requirement to use third-party technology for storage and processing of educational literature. In this case, the approach will allow you to quickly move from a monolith to a microservice architecture.

During development, a monolith with observance of modularity will be created, which will allow to extract the module responsible for processing of educational literature without unnecessary efforts.

Development

Each of the software tools or approaches was involved based on a specific architecture. Given that the architecture in the system clearly separates the server part from the client, both parts were considered separately.

No less important is the protection of the web-system. In order for the developed system to have good protection, a number of approaches and software were used such as JSON Web Tokens or Spring Security.

Testing is an integral part of the software development process. Moreover, one of the conditions of a certain architecture is test coverage,

because it helps to monitor the quality of written code. There were tests of different types, including:

1. modular testing of the client and server part of the system
2. integration testing of the server part of the system
3. architectural testing

An information system uses several data storage solutions at once. Because learning resources must be stored separately from other data, different data warehouses have been involved and separate services have been developed to process the information before it reaches the repository itself.

The already developed system was deployed on the Heroku cloud platform.

Conclusions

In the course of the work, a system was created that automates the process of distribution of educational literature, making it easy to use and easy to implement.

The developed system relieves the teacher of responsibility in managing the distribution of educational materials. The system takes on the task of sending students books, notes, laboratory work and other resources on the exact day and time. The system improves the quality and speed of students' mastery of educational materials, compared to traditional approaches. Thanks to the single structure of e-mails used by the system, each student can filter e-mails as he wishes and find the materials he needs.

The possibility of using non-standard data storage solutions by the educational institution was also taken into account. This became one of the decisive arguments in choosing the approach to building the architecture of the information system.

A new approach to database development has been mastered, version control of database schema changes. This simplified the process of building a database throughout the development of the web-system.

The developed information system is easily implemented due to the fact that the student does not have to have an account in the system, it is enough to have mail, which will receive all the necessary information with content partially generated by the system and partly by the teacher.

References

1. Bereza A.M. Fundamentals of information systems. 2 editions, revised and supplemented – K.: KNEU, 2001. – 204 p.
2. Fowler M. Monolith First [Electronic resource] / Martin Fowler. – 2015. – Access mode: <https://martinfowler.com/bliki/MonolithFirst.html#footnote-typical-monolith> (last access: 26.05.2020).

IMPLEMENTATION OF PROCEDURES FOR BUILDING WIRE-FRAME STEREOSCOPIC IMAGES WITH MULTI-ANGLE OBSERVATION

Introduction

Frame 3D graphics are an important class in a range of other spatial imaging tasks. On the display screen, frame computer three-dimensional images are represented by a set of points and / or vectors.

There are many software solutions for working with 3D graphics. Many of them also implement work with stereoscopic images (Blender, Cinema4D, 3DSMax, etc.). However, these programs do not implement multi-angle observation. Most of the programs that have implemented this task are narrowly specialized and are not publicly available. Examples of such software solutions are X-ray scanners of luggage at airports, three-dimensional scanning of objects, terrain, etc. [1]. However, most solutions only use multi-angle as a tool to obtain information about the spatial coordinates of points on the object to further create its three-dimensional model and do not offer stereo output to the monitor.

The goal of the research is to develop an application capable of modeling 3d objects with stereoscopic output with reacting to the relative position of the user to the screen that is input via camera.

Solution

The principle of construction of stereoscopic projections for graphic three-dimensional images is well known [2].

In IPME after G.Ye.Pukhov NAS of Ukraine proposed and developed a new operator method of stereo transformations [3,4,5], the essence of which is to find a formal mathematical apparatus that establishes a one-to-one correspondence between the three-dimensional coordinates of an arbitrary point given in the world coordinate system, and the stereo coordinates of this point specified in the screen coordinate system. The transition from the spatial area to the stereo area is performed using the so-called direct stereo conversion operator:

$S\{\vec{V}\} \Rightarrow \vec{s}$. The reverse transition is performed using the inverse stereo conversion operator (Fig. 1).

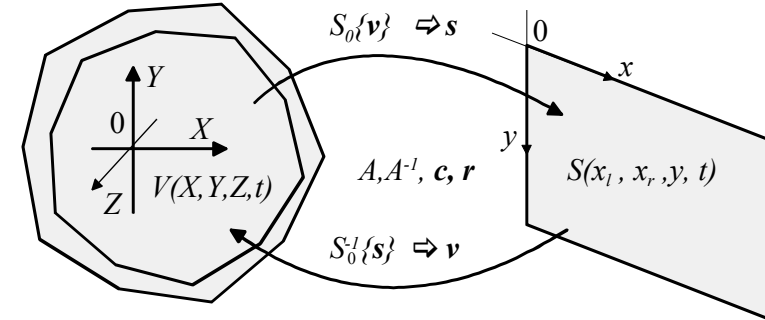


Fig. 1 - operator method of stereo transformations

This approach made it possible to solve the problem of connected (corresponding) points for graphic computer stereo images. Matrix dependences reflecting this transformation are obtained for the case when the observer is at the point of a constant stereo angle.

Direct and reverse stereo operators of a constant foreshortening are realized by means of the following dependences:

$$\vec{s} = A(\vec{V} + \vec{c}) / (z_0 - Z - \dot{z}) \quad (1)$$

$$\vec{V} = (z_0 - Z - \dot{z}) / A^{-1} \vec{s} - \vec{c} \quad (2)$$

$$A = \begin{bmatrix} z_0 & 0 & a - x_0 \\ z_0 & 0 & -a - x_0 \\ 0 & -z_0 & -y_0 \end{bmatrix}, \quad A^{-1} = \frac{1}{2az_0} \begin{bmatrix} (a + x_0) & (a - x_0) & 0 \\ -y_0 & y_0 & -2a \\ z_0 & -z_0 & 0 \end{bmatrix},$$

$$\vec{V} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad \vec{c} = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix}, \quad \vec{r} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}, \quad \vec{s} = \begin{bmatrix} x_l \\ x_r \\ y \end{bmatrix},$$

where $2a$ – stereobase, A, A^{-1} , – forward and inverse square matrices,

\vec{V} – vector of three-dimensional coordinates of an arbitrary point of the object,

\vec{s} – vector stereo coordinates of the desired point,

\vec{c} – vector of offset of the origin of the world and screen systems,

\vec{r} – the vector of the constant point of view.

Object recognition using a cascading Haar classification is an effective method of object recognition proposed by Paul Viola and Michael Jones in their article on methods of rapid pattern recognition [6]. A machine learning approach was used, in which the cascade function is trained from a

large number of positive to negative images. What is next used to recognize objects in other images.

The implementation of multi-angle results in the need to calculate the observation angle of each frame, which, provided the formula (1) leads to the need to calculate the original coordinates of the stereo pair by formula (2), which is a resource-intensive process. To simplify the calculations, it is proposed to move from the "object-centric" to the "camera-centric" model. Then the formula for obtaining screen coordinates will look like this:

$$\vec{s} = \begin{bmatrix} (x'_0 X - (a - x_0)Z)/d \\ (x'_0 X - (a + x_0)Z)/d \\ z'_0(-Y) + y'_0 Z/d \end{bmatrix}, \quad \vec{r}' = \begin{bmatrix} x'_0 \\ y'_0 \\ z'_0 \end{bmatrix} = \vec{r} + \vec{c}, \quad d = z'_0 - Z,$$

where \vec{r}' – the transformed vector of the constant stereo angle,

\vec{r} – predefined stereo angle or initial camera coordinates,

\vec{c} – coordinates of the observer in front of the monitor or the modifier of the camera coordinates.

With this approach, all operations on objects take place in a "traditional" three-dimensional space, which reduces the number of calculations required for the various tools implemented in this work, reduces the amount of memory used and increases the simplicity and reliability of the source code.

Results

As a result an application for 3d modeling was created. So far it has a rather limited functionality, such as drawing polylines, extruding them into meshed objects, and basic transformations – translation, scaling (Fig. 2), rotation of these objects. For rendering graphics was used OpenGL library. For detecting user position was used OpenCV library (Fig. 3). The position is detected in 3 dimensions although the accuracy of position detection has a room for improvement. For example the detection model can be further trained and received data can be better interpolated.

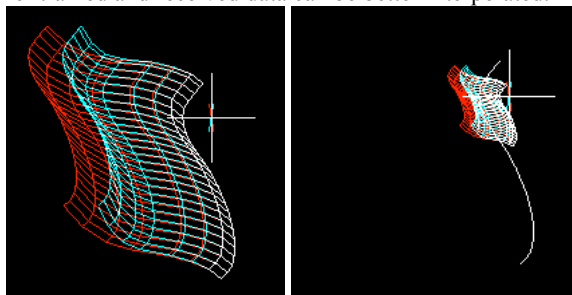


Fig. 2 – Scaling object

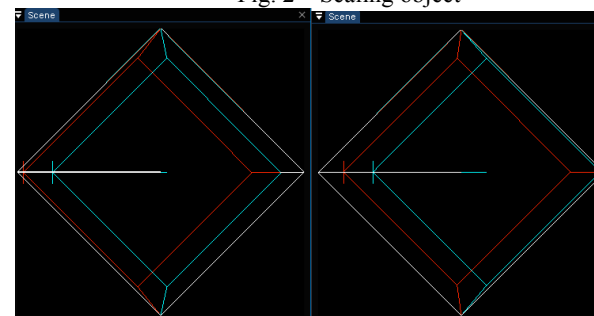


Fig. 3 – change of the output image based on user position

References

1. Daneshmand M. 3D Scanning: A Comprehensive Survey / M. Daneshmand, A. Helmi, E. Avots, F. Noroozi, F. Alisanoglu, H. S. Arslan, J. Gobrova, R. E. Haamer, C. Ozcinar, G. Anbarjafari // arXiv — 1801.08863v1, 24 Jan 2018, 18 p.
2. Valus N.A. Stereoscopy. — M. : Hayka, 1962. — 378 c.
3. Grutz Yu.N. Stereoscopic computer graphics. — Kyiv: Scientific Thought, 1989. — 160 pp.
4. Gruts Yu.N. Stereoscopic operators and its application / Gruts Yu.N., Jung-Young Son // Proceedings of The 6th International Workshop on 3-D Imaging Media Technology and The 5th Photonic Information Processing Conference. — 2000. — Vol.6, N1, pp. 34-38.
5. Grutz Yu.N. Operator method of stereo transformations in computer graphics / Proceedings of the 6th International Scientific and Technical Conference: Information Systems and Technologies IST-2017, dedicated to the 80th anniversary of V.V. Sviridova September 11–16, 2017. - Koblevo, Ukraine. - P. 37-38.
6. Viola P. Rapid Object Detection using a Boosted Cascade of Simple Features / P. Viola, M. Jones // Accepted conference on computer vision and pattern recognition, 2001 –9 p.

MOBILE SENSORS IN INTELLIGENT TRANSPORTATION SYSTEMS

Over the last few years, smartphone-based data collection has become an important complementary technique for detecting road anomalies in order to monitor the surface of road sections or the surface of bicycle and pedestrian lanes. Sensors are expected to be a new area in which smartphone-based measurements seem particularly attractive because they are widely used.

Measurement and analysis of motion sensors from different types of mobile devices can vary depending on many factors, including the characteristics and quality of the sensors, the position of the smartphone device, the car's suspension system and speed.

Measuring signals from smartphone sensors is a difficult task due to the different properties of the sensors in different smartphone models, differences in their size, weight, length and suspension systems of the car. There is also a difference in the movement of the vehicle caused by road anomalies, namely: the length, depth and shape of potholes on the road surface and the effect of curvature of the road on the speed of different vehicles [1].

Poor road conditions can cause damage to vehicles, increase maintenance costs and fuel consumption, reduce driving comfort and may even increase the risk of accidents.

Different vehicles passing through a particular pothole do not create an identical signal pattern. In addition, the speed of the vehicle affects the speed of change of motion sensors, which leads to different models of vibration response to any anomaly of the road surface.

Modern smartphones have a large number of types of built-in sensors. Some of them are hardware (physical), and others - software (virtual) [2].

Hardware sensors are physically embedded sensors, such as an accelerometer, gyroscope, magnetometer, light sensor, temperature, and so on. And software sensors use data from one or more hardware sensors and calculate values in real time, for example: linear acceleration, rotation, gravity and others.

Smartphone sensors can also be classified into three different types: motion sensors, position sensors, and environmental sensors.

Motion sensors are suitable for monitoring the movement and vibration of the device, tilt, vibration, rotation or rocking. Gestures can

directly reflect user interactions with an application, such as when a user controls a car or controls an object in a mobile game.

Position sensors are suitable for determining the physical position of the device in the local coordinate system. For example, a geomagnetic field sensor in combination with accelerometer sensors can determine the position of the device relative to the local coordinate system.

Environmental sensors measure the environmental properties of the surrounding area, such as temperature, humidity, atmospheric pressure and light [2].

For each smartphone-based application, different combinations of sensors can be used depending on the desired application criteria. With the help of these sensors it is possible to develop an application for pavement analysis, detection of anomalies.

The motion sensors can be used to monitor the oscillation or tilt of a moving vehicle caused by road anomalies.

GPS (Global Positioning System) and Wi-Fi can be used to determine the current location of the smartphone (latitude, longitude), direction orientation and speed.

There are three main approaches to road condition monitoring:

- based on 3D reconstruction
- based on video surveillance
- based on car vibrations

The 3D approach relies on laser 3D scanning to create accurate digital surface models. That is, 3D scanning is the process of analysing an object or environment in the real world to collect data about its shape and appearance. The collected data can then be used to build digital 3D models.

The 3D scanner can run on many different technologies, each with its own limitations, advantages and costs. Of the significant disadvantages, optical technology can face many difficulties, such as shiny or transparent objects.

In this approach, the laser 3D scanner uses reflected laser pulses that create accurate 3D-digital models of existing objects, such as road anomalies. Due to the accurate model and visualization of the terrain, there are many opportunities for the best analysis of the road. However, the above approach requires many laser scanners and requires significant costs when monitoring large-scale road networks. [3].

The approach based on video surveillance relies on the analysis of image processing of video cameras, on the basis of which the road texture is found and compared with already taken photos of certain road anomalies. This system uses terrain-captured images taken by a camera or video system mounted on a moving vehicle. Any suspicious objects on the road surface, including potholes and cracks, can be automatically detected from the

collected video images. Although this approach is more economical than the first, which is based on 3D reconstruction, it depends on certain environmental conditions, such as lighting, shadows, etc. [4].

Using a vibration-based approach, road events are detected by the speed of vibration of moving vehicles captured by motion sensors (eg accelerometers or gyroscopes). Theoretically, the vehicle when vibrating through any anomaly of the road surface, such as potholes, cracks, hatches, will vibrate more than when passing on smooth road surfaces [5].

At the stage of processing, pre-assembled, analysed and reduced to a user-friendly value of the sensors are used to detect road anomalies. The values of the accelerometer are most favourable for the analysis of road anomalies among other sensors, by tracking the acceleration of the object. When moving on a rough road, the car creates slight oscillations in space, making it possible to detect abnormal changes in the sensor. There are three basic approaches to this analysis:

The first approach is based on the comparison of pre-experimentally determined and analysed threshold values of sensor indicators, with new current values.

The second approach is based on machine learning. It uses more advanced techniques and algorithms to detect pavement. For example, it is possible to use the reference vector method, which is a method of analysing data for classification using managed learning models. For a given set of training samples, which were assigned to one or another category (good or bad road surface), the algorithm builds a model that assigns new samples to one or another category. The model of this method is the representation of indicators as points in space reflected in such a way that they are separated from individual categories by a hyper plane with the maximum interval in this space [6].

The third approach is based on the algorithm of dynamic transformation of the time scale (DTW-algorithm). This algorithm allows to find a correspondence between two time series and is mainly used in speech recognition research. The algorithm compares the input data of the signals with predefined patterns and measures the similarity between the data sets.

Patterns of anomalies in digital signals occur when the signal strength exceeds a certain value. As a basis, to calculate the threshold values, three issues were considered: the duration of the interval for the window function fixed against the flexible definition of the threshold and the amplitude of the signal relative to other properties of the signal amplitude (mean and standard deviation).

Determining the length of the interval for the window function in spectral analysis is a difficult task because it is related to various factors, such as the speed of vehicles and the distance from the front to the rear

wheel. The window function takes into account predefined signal data intervals for analysing and retrieving functions as opposed to viewing signal data separately.

Determining the correct thresholds in the statistical approach is an intensive process, as the value of road anomaly data is affected by different conditions. The car's suspension system, the touch properties of smartphones and the placement of smartphones all affect how smartphones experience the same anomaly.

The accelerometer has a three-dimensional Cartesian coordinate system relative to itself, represented by orthogonal axes x, y, z. In addition, we determine the Cartesian frame of reference relative to the vehicle in which the accelerometer is located.

The problem of detecting road anomalies is difficult for several reasons. The main task is to establish the correct and true threshold value for further analysis. This is usually done only with a manual annotation, which is subjective and can vary considerably depending on how the vehicle approaches the anomaly and at what speed. However, the amplitude of the signal may have different characteristics at different speeds.

One of the most important indicators for the further use of the obtained information is the binding to the area. After research, it was found that the allowable deviation for further analysis is 3.5 - 6 meters. With a larger deviation, it will be difficult to work with the information attached to this location. Under normal conditions, the deviation averages 4 meters, which is sufficient for further analysis and corresponds to the typical measurement errors of modern GPS receivers on the street.

One of the most significant disadvantages of a smartphone-based system is the power consumption of the device and its energy consumption.

Several experiments were conducted to determine the most energy-intensive processes. The device's power consumption experiments were performed on an iPhone XS smartphone using the xCode Energy Impact and CPU Instruments. The most energy-consuming processes were experimentally established:

- frequency of updating of motion sensors
- calculating operations for filtering and data processing before sending to the server
- GPS accuracy

The interval of updating of motion sensors was considered in the range of 0.05 - 0.2, at higher frequency of updating too big value of power consumption of the device was revealed.

Summing everything up it is important to note that mobile sensors can be widely used in intelligent transportation systems for different purposes. Firstly, for detecting road condition and creating the best route

using not only the traffic information, but also the information about road condition.

References

1. Власов В. Транспортная телематика в дорожной отрасли / В. Власов, В. Богумил, Д. Ефименко. – М.: МАДИ, 2013. – 80 с.
2. Labrador, M. Human Activity Recognition: Using Wearable Sensors and Smartphones / M. Labrador,, O. Lara Yejas., 2013. – 207 с.
3. Hou Z. Experimentation of 3D pavement imaging through stereovision. In Proceedings of the International Conference on Transportation Engineering / Z. Hou, K. Wang, W. Gong. – Chengdu, China, 2007. – 376 с.
4. Yagi K. Extensional smartphone probe for road bump detection / K. Yagi. – Busan, Korea, 2010.
5. Kim T. Review and analysis of pothole detection methods. / T. Kim, S. Ryu., 2014. – 603 с.
6. A road pavement monitoring system for anomaly detection using smart phones. In Big Data Analytics in the Social and Ubiquitous Context / F. Seraj, B. van der Zwaag, A. Dilo, T. Luarasi. – Berlin, Germany, 2014. – 128 с.

Karaieva N., Varava I.

METHODOLOGY DESIGN OF MONITORING SYSTEM OF INDICATORS OF THE HUMAN DEVELOPMENT LEVEL OF UKRAINE'S REGIONS

As a tool for measuring and comparing the socio-economic achievements of many countries in the early 1990s, experts from the United Nations Development Program (UNDP) had proposed to calculate the Human Development Index (HDI). As noted in the UNDP Report "Human Development Indices and Indicators: 2018 Statistical Update" [1], developed HDI embodies the progress of humanity, bringing together in one indicator information about health, education, and income of people. Since 1990, the results of corresponding calculations have been published in international annual reports on the HD and are used for monitoring and ranking assessments of the state of many countries around the world. For Ukraine, the calculations of HDI were first included in the UNDP report on human development in 1993.

In world practice widely using information-analytical systems (IAS) for HD monitoring, that was implemented on basis of UNDP Report "Human Development Indices and Indicators: 2018 Statistical update" and presented as statistical Web application "Human Development Indices and Indicators: 2018 Statistical update". The site presents monitoring and analysis (for the period 1990-2018) of constituent indices and indicators of HD for 189 countries (including Ukraine) and territories. (fig. 1).

Visualization of analytical data processing results for each country and territory is presented in the form of tables, graphs, and profiles, which characterize a position (rating) of a certain country in a group of other countries by total HDI. Moreover, the user is provided with an overview of the analytical note text file for each country, which contains country characteristics by key human development indices, in particular, Human Development Index (HDI), Inequality-adjusted Human Development Index (IHDI), Gender Development Index (GDI), Gender Inequality Index (GII) and the section on five information panels.

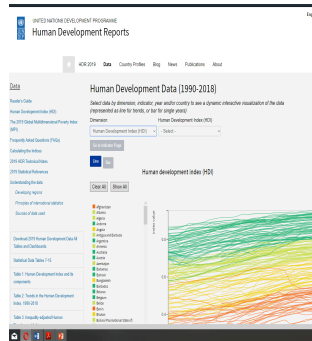


Figure 1. The main page of Web application "Human Development Indices and Indicators: 2018 Statistical update"

Using the world experience, the basis for the formation of the IAS monitoring of the regional HD of Ukraine is the development of such structural blocks (or subsystem), as (fig. 2): thematic multidimensional database (DB) and knowledge base (KB); analytical block; visualization block.

Thematic DB and KB should be represented object-relation normalized database tables, which containing a hierarchical system of interrelated indicators of the level of HD and environment state. In general, regional HDI calculations performed by the two-level scheme [2]:

- 1st level: construction of indexes by different indicator blocks (comfortable life; welfare; decent work; education);

- 2nd level: construction of the integral index. 33 indicators are included in the calculation of the regional HDI. The integral indicator of the environmental state is calculated separately, which determinate by indicators of air state, land and water resources due to anthropogenic load and natural state of regions ecosystems of Ukraine.

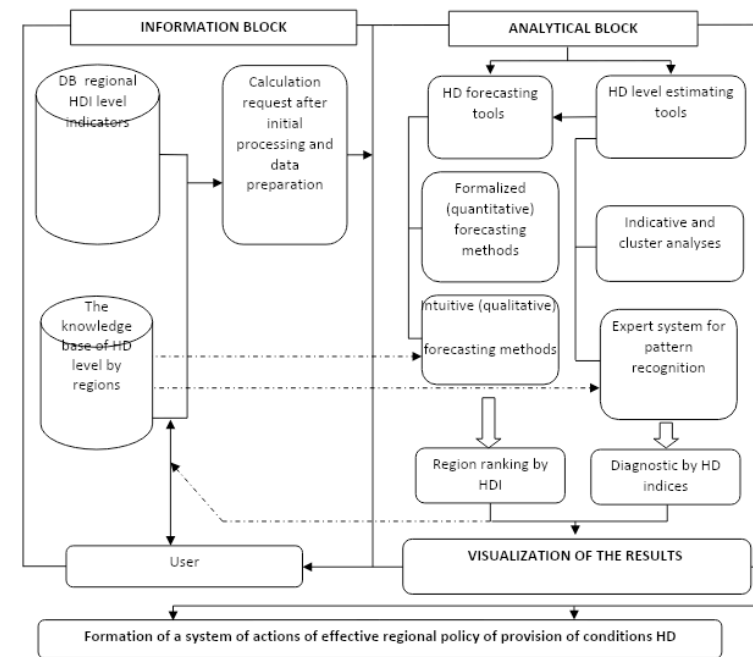


Figure 2. Structure of IAS for monitoring of HD indicators of regions

Annual national statistical reporting is the information base of calculations. The knowledge base accumulates theoretical and knowledge of experts about the object of study in the form of a classis's description. The essence of the DB and KB concept lies in integrated storage and differentiated using by a software of whole information about domain objects that provide a certain interest to the user. The information which describes knowledge about the system and the situation in a crisis condition is most adequately formed using the methods of fuzzy logic theory. To ensure the formation of user queries, it is necessary to develop a scientific and methodological system that contains calculation methods, regulatory and legal documentation, directories, profile archives, analytical notes, publications on the subject of HD.

MS SQL Server is suggested to use as a database with MS Analysis Services and MS Reporting Services technologies.

The **analytical block** should satisfy the user's needs for obtaining

analytical information depending on the type of applied applications. The main components of analytical information processing are dynamics of change of regional indicators, HD indexes; ranking of regions by HDI; forecasting of key HD indicators; region ranking (grouping) by HD indices on basis of cluster analysis; diagnostics of the regions state by threats level (or prerequisites) of HD.

The development of an analytical block relates to some difficulties given a significant number of forecasting methods. The main mathematical models, that used to predict different indicators are correlation and regression models, models based on the principal component analysis, neural net models. One of the prerequisites for the construction of predictive statistical models of socio-economic development of the country is the identification of homogeneous sets of regions represented by the system of HD indicators. Cluster analysis is an effective method that allows to group districts into homogeneous groups using a wide range of indicators. Cluster analysis results allow to determinate disproportions of regions development by investigate indices and developing effective State policy on this basis.

At present, the use of Statistica and SPSS software packages greatly accelerates the cluster analysis process itself. But, in our opinion, an important stage of cluster analysis with these packages is the form of presentation and interpretation of the results.

The **visualization block** provides the presentation of regional information in tabular, text, graphic and cartographic form. In the subsystem must be realized functional opportunities of analysis results presentation and calculated regional HDI values and integral indicator of environment state with using business graphic tools, export to MS Excel and other formats, and also printing out information. Important elements of the visualization block are the formation of a single profile template (passport) for each region and a short analytical note (single template) for the specified period. The data, that characterizing the score of a certain region among other regions by overall HDI and its calculated indices and components of integral index of environment state, should be presented in the profile. Presenting of dynamics of changes of indices vales is advisable to assess the effectiveness of regional policies, for example, in the past three years. Dynamics of rating and values of HDI components changes allows describing in more detail trends of the formation of threats to human development in the region.

The *software of IAS* should be based on the use of modern data warehouse concept, data mining techniques, online analytical processing (OLAP), network technologies of information services for users. The system is proposed to be implemented as an ASP.NET Core MVC application and deployed in Amazon Web Services (AWS). The client-side

is implemented using the Angular framework with the use of Leaflet Open Map Library. The system will be accessible for end-users through any web browser: Internet Explorer, Mozilla Firefox, Opera, Safari, etc.

Conclusion. Overall, the studies show that Expenditure planning for regional budgets should be based on HD indicators under the policy of decentralization. Implementation of the IAS of development indices of Ukraine's regions can contribute to planning effective regional policy activities and democratic transformations in the regions by ensuring access for all to information resources and provision of information services by local governments using the Internet.

References

1. Human Development Reports UNDP "Human Development Indices and Indicators: 2018 Statistical Update". URL: <http://hdr.undp.org/en/2018-update> (дата звернення: 10.03.2020).

2. Методика вимірювання регіонального людського розвитку; затверджена Рішенням Президії НАН України та колегії Державної служби статистики України від 13.06.2012 р. № 123-м. URL: HTTP://WWW.IDSS.ORG.UA/ARHIV/METODIKA_RLR.ZIP (ДАТА ЗВЕРНЕННЯ: 10.03.2020).

THE USE OF IMAGES FOR STEGANOGRAPHIC INFORMATION PROTECTION

The problem of protecting information from abduction and unauthorized use has always existed. On the one hand, in our time with the emergence, development and widespread use of information technology, there are new risks of violating the confidentiality of information; on the other hand, new methods and approaches to solving this problem were appeared. The threat of theft of important information exists in case it is stored in computer systems, on external storage media and while it is transmitted over a network. Software protection involves the use of passwords, electronic digital signatures, cryptographic and steganographic methods. Steganographic methods [1] are quite reliable and deserve special attention.

The use of steganographic method of protection has the advantage of hiding the fact of transmission of classified or confidential information reduces the risk of falling into the wrong hands.

There are five main areas of steganography, that are most often allocated in the literature: classical, which involves the use of chemical or physical properties of various information carrier, linguistic method which use conditional letters and semagrams, digital method, in which the properties of computer formats are used for hidden information, quantum method, which in practice has just begun to develop and is based on the laws of quantum physics and classical information theory; network method, in which features of protocols work of data transmission by computer networks are used for information hiding [2].

At the present stage of information technology development, digital computer steganography is most widely used. Its methods are based on embedding secret messages (text, images and other files) into digital data, usually images (frequently .jpeg, .gif, .bmp formats), video (uncompressed .avi-file is a sequence of .bmp-images) and audio recordings; quite rarely the information is hiding in a texts.

Containers for hidden messages are most often used because they contain a lot of redundant information that is not perceived by the human eye. In the three-channel RGB model, different colors correspond to different intensities of the three basic colors (RGB — red, green, blue). In this case, if the color information of each image pixel is encoded with 24 bits, then for each base color 8 bits are used and the value of the intensity of each component color can be a number from 0 to 255, which allows to encode almost 17 million different colors. When encoding a color with 36

bits, 12 bits are allocated for each channel (gradations are from 0 to 4095; therefore, you can encode more than 68.7 billion colors). There may also be 8-bit encoding, but it only gives 256 different colors. The choice of encoding model affects to the image size.

The more data you need to record, the larger an image container size should be supplied. It should be borne in mind, that for storing images in files, different formats are used, which involve the use of different methods of data compression. Today, about 80% of images are encoded with a 24-bit model. Therefore, you need to choose a 24-bit model for encoding images to successfully hide the information, and, in addition, the image should be as large as possible, because the higher the image resolution is, the less the neighboring pixels will differ. It is recommended to use .bmp-files as a container as non-compressible, (because when the image is used as a container, the file size does not increase) [3], although this format is rarely used now, and many methods have been developed to detect hidden information; or the most common are .jpeg image files for storing digital photos, in which image compression is best implemented with minimal loss of visual quality; either one of the most common formats on the Internet — gif, which uses a lossless compression algorithm.

The algorithm for hiding information in the image is based on changing certain bits of a single pixel, so that the distortion is lower than the level of recognition by the human eye. As a result, there are many different implementations, that differ by which bit to choose or in what sequence this should be done etc.

As mentioned above, each base color of the three-channel RGB model in one pixel of a 24-bit image is encoded by one byte (8 bits), the text character of the *char* type also occupies one byte. But if you replace the byte of the image with a byte of text, you will get a completely different color and, as a result, the image would be spoiled. Therefore, the best option is to replace one bit of any of the channels or all channels simultaneously. In this case, it is important which bit of eight bits will be replaced. Often the last bit or the last 2-3 bits should be chosen. For example, let the byte that corresponds to one of the channels, store the $11111111_2=255_{10}$. Then, if the first bit is replaced by 0, we get the number $01111111_2=127_{10}$, so the intensity of the color component will be approximately twice less, which will cause noticeable to the human eye color changes of the pixel image (for example, white color with such a change in the green component become pink). If you replace the last bit, you will get the number $11111110_2=254_{10}$ (if you replace the penultimate bit, you will get $11111101_2=253_{10}$, or in case the third from the end — $11111011_2=251_{10}$, or even the last three, then $11111000_2=248_{10}$). That is, the intensity of the component color will change insignificantly and the human eye will not notice this in the image.

In most cases, one last bit is chosen to hide information. This is due to the use of sampling error that exists in digitized images, videos and audio files [3]. However, to hide as much information as possible, the last 2-3 bits of each pixel color channel are also used [3; 4], for example, 2 red, and 3 both green and blue, which makes it possible to hide 8 bits of information in 24 bits, and such changes will not be noticeable to the human eye. Changes will be less noticeable if the hidden information is placed on the edges of the image. In addition, the smaller the hidden message is, the harder it is to detect. Sometimes only the blue color channel is used to hide information, because the human eye is the least sensitive to blue.

Using only one bit allows you to hide, for example, information of 100 KB in a container of 800 KB, and when using 2-3 bits (4 of 24), you can hide information of 100 KB in a container of 300 KB.

Although the human eye does not perceive minor changes in color, but there are methods that make possible detecting hidden information. Steganoanalysis methods are used to prevent and detect hidden information. Thus, the method based on the statistical criterion of χ^2 -Squared Test makes it possible to detect infiltration of information by replacing the smallest significant bit of the image, but it does not work with pseudo-random selection of lowest bit. This is happen, because if the last bit was 0, then after inserting the information, its value is either remain 0 or become 1, so it will not decrease, and, conversely, if the value of the last bit was 1, it will not increase. If build a pixel histogram of an image, then for an empty container the height of individual columns of intensities will change smoothly and for a filled container this change of height of individual columns will not be smooth, so the smoothness of the histogram is violated. Therefore, it is better to use "noisy" images obtained from a scanner or camera to hide information, rather than created on a computer by graphic programs. The Sample Pairs Analysis method [5] makes it possible to estimate the number of hidden bits in the image. One of the most accurate modern method of detecting hidden data in images is RS-stegoanalysis [6], which is based on statistical and combinatorial methods and examines the relationship of the least significant bits of data with neighboring bits.

Analysis of various implementations of steganographic coding shows that almost all algorithms use additional cryptographic data encryption, which increases the guarantee that hidden information, even when exposed, cannot be read quickly and easily.

There may be one or more stegokeys in the stegosystem, depending on the degree of protection (for example, a pre-encrypted message is inserted). In steganography, also as in cryptography, there are two types of keys: private and public. The stegosystem with a private key uses one key, which must be declared either before the beginning of the exchange of a secret message, or

transmitted through a secure channel. The public key system uses different keys, which differ in such a way that it is impossible to derive one key from another using calculations. Therefore, one key can be transmitted freely through an unsecured communication channel.

Simple algorithm was developed and programmed in C++ under the Microsoft .NET platform to perform steganographic encoding of text information, also for encrypting and decrypting characters (without the use of cryptography) in a 24-bit image of .jpeg, .png, .bmp formats. When encrypting, character by character is selected from the text, and individual bits are sequentially allocated from the characters. Simultaneously, information about an individual pixel is read from the image file, in one of the channels, the last bit is replaced by a selected bit and the changed information about the pixel is written back to the file. Decryption is implemented in a similar way. The key is the pixel number of the image, from which the hidden information is stored.

If you add cryptographic methods to this algorithm, for example, before writing a message to an image, encrypting it using the AES-256 method with a 256-bit key or similar, the protection of information will be reliable.

References

1. Gary C. Kessler. An Overview of Steganography for the Computer Forensics Examiner. — garykessler.net/library/fsc_stego.html
2. Весельська О.В., Зюбіна Р.В., Фролов О.В. Систематизація та класифікація наявних стеганографічних методів приховування інформації // Наукоємні технології. — № 2 (30), 2016. — С. 187-194.
3. Денисюк В.О. Реалізація стеганографічного алгоритму захисту даних з використанням файлів зображень // Техніка, енергетика, транспорт АПК. — № 1 (100), 2018. — С. 29-38.
4. Шадхін В.Ю., Грищенко О.Л., Голосков А.С., Хижняк Д.С. Дослідження методу LSB для стеганографічного приховування інформації у вигляді зображень // Технології та дизайн — № 2 (3), 2012. — С. 1-7.
5. Швідченко І.В. Про один метод стеганографічного аналізу контейнерів-зображень // Штучний інтелект, 2013, № 3. — С. 554-563.
6. КОРОЛЬОВ В.Ю., ПОЛІНОВСЬКИЙ В.В., ГЕРАСИМЕНКО В.А., ГОРІНШТЕЙН М.Л. КОМПЛЕКС СТАТИСТИЧНИХ ДОСЛІДЖЕНЬ ДЛЯ СТЕГОАНАЛІЗУ // МАТЕМАТИЧНЕ ТА КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ СЕРІЯ: ТЕХНІЧНІ НАУКИ. ВИПУСК 5, 2011. — С. 134-149.

SOFTWARE TOOLS FOR INTELLECTUAL AND STATISTICAL PROCESSING OF HYDROACOUSTIC SIGNALS

Hydroacoustic studies the phenomena that occurs in the aquatic environments and associated with the formation, propagation and reception of acoustic waves [1]. But despite the large number of sonar devices, hydroacoustic still faces big problems with data accuracy.

Collected from sonar devices data can be processed and analyzed in different ways. One of the promising areas is usage of machine learning [2]. Creation of artificial conditions is a great possibility to generate required amount of digital information. With this information it will be possible to create machine learning models that will be used for further data forecasting or anomalies detection in aquatic environments.

With statistical processing and modeling, you can get more accurate data about the location, size, speed, nature of objects. Machine learning can also provide opportunity to create systems for detection and elimination of extraneous noise in data obtained from sonar devices. That kind of information can be used in the military, maritime navigation, fish exploration, oceanographic research, etc.

The system for processing sonar signals is implemented using Python programming language [3]. The analysis of sonar signals is primarily reduced to the processing of information from audio tracks, presented in .wav format. Files of this format have a conform structure that includes meta-information part and the audio data itself. This structure is shown in table 1.

Table 1. Audio track data structure

Cell size	Cell name	Description
4	Chunk ID	“RIFF” chunk descriptor. The Format concerns
4	Chunk size	
4	Format	
4	Subchunk1 ID	“fmt” sub-chunk
4	Subchunk1 size	
2	Audio format	
2	Num channels	
4	Sample rate	
4	Byte rate	
2	Block align	
2	Bits per sample	
4	Subchunk2 ID	
4	Subchunk2 size	

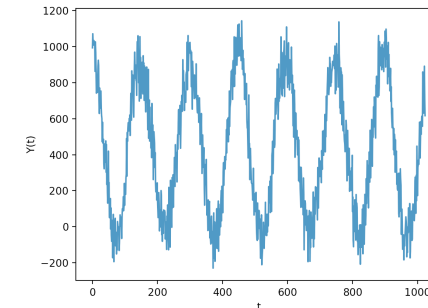
*	Data	“data” sub-chunk
---	------	------------------

Initial step of dataset creation is processing of audio track’s meta-information. Then, based from information received, it becomes possible to read raw audio data, recorded in the .wav-file. This step can be simplified by usage of Python package called wave, it contains predefined functions for .wav-files processing, but in that case some additional flexibility in data-processing can be lost. If usage of this library does not meet certain requirements, the built-in Python struct package can be used, which allows iteratively walk through the data stream of the file and read the bytes in the specified order.

One more approach is usage of ready-made dataset created by the special sensors. Those data will be partially similar to the data, gathered from audio tracks, but may contain more accurate and additional values. The special case of extended data is data that consists not only from the values of the pressure at a certain point (p), but also the displacement of the particles of the medium along the coordinate axes (v_x, v_y, v_z). Such data is more accurate due to the possibility to work with the same signal in several dimensions. Filtration, noise reduction, values extraction, etc. will be much more efficient with that kind of data.

For forecasting execution, samples of relevant data (a set of data that is relevant to audio signals and provides metrics, that needs to be predicted) required along with audio data. Anomalies detection does not require any additional information, this method focuses on searching of instabilities in the input data.

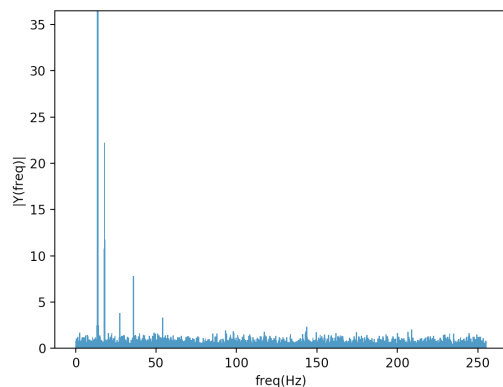
From the signal data obtained, it is possible to build an oscillogram, which can visually, to some extent, show some information about the signal, but that kind of data will not be effective on machine learning model creation without pre-processing. An example of an oscillogram is given on picture 1.



Picture 1 - Oscillogram for 0-0.5 s.

In the case of signal analysis, the spectrum will be effective for

analysis. Since sonar signals frequencies are quite low, it makes sense to limit them to range of [0; 255]. Picture 2 shows the spectrum of the same signal was given on the oscillogram.

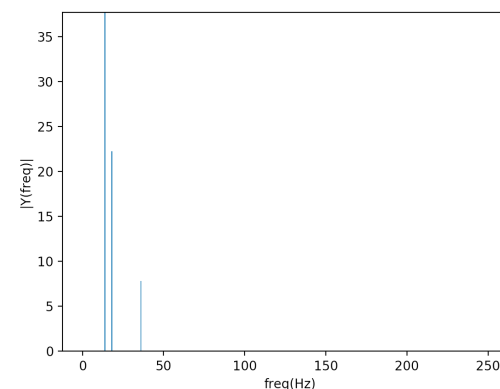


Picture 2 - Signal spectrum.

Signal, described by spectrum, will be much more efficient for model creation. Since sonar signal spectrum is limited by specified range, we can send those data to machine learning model as vector, sized by length of specified range, in our case size equals to 256. But still data efficiency can be increased by noise reduction. One of the possible solutions to reduce noises is usage of standard deviation, but in reverse notation: abnormal values should be kept, normal values should be filtered out. Picture 3 shows example of signal spectrum with filtered noises.

After filtration execution, the data will become effective for creation of machine learning model.

Information from audio tracks is translated into a tabular view and stored within the database which is related to the implemented system. The database also stores project configuration and output data. The PostgreSQL relational system was chosen as the database management system; program interacts with database by usage of python package called psycopg2 in combination with ORM-component SQLAlchemy. Matplotlib is used to display information from audio tracks and resulting graphs. Thanks to the visual presentation of information, you can easily see the relationship between input and output data.



Picture 3 - Filtered signal spectrum.

User interface is created using PyQt5 library modules. Data processing and calculation are performed by the functions of libraries NumPy (work with multidimensional arrays), SciPy (processing of numerical sequences), Pandas (work with two-dimensional tables). The scikit-learn machine learning library was used to build the models. The project code configures the interaction between the selected packages and connects all the components into a single system.

Implemented software product can provide information about objects and detect anomalies in the aquatic environment.

References

1. Румынская И.А. Основы гидроакустики. — Л.: Судостроение, 1979. — 209 с.
2. Вьюгин В.В. Математические основы машинного обучения и прогнозирования. — М.: МЦНМО, 2018. — 384 с.
3. PYTHON 3.8.3 DOCUMENTATION [ELECTRONIC RESOURCE]. — ACCESS MODE: [HTTPS://DOCS.PYTHON.ORG/3/](https://docs.python.org/3/)

Kuzminykh V., Kotsubanov G.

THE DECISION-SUPPORT PLATFORM FOR INTERNATIONAL SCIENTIFIC AND TECHNICAL COOPERATION

Immersed in the concept of scientific and technological development, the most useful for the civil good is the concept of coherence of actions of knowledge-producing countries. There are many international organizations that bring together many countries from all over the world in certain fields of science. Accordingly, the interaction process should be coordinated and adjusted in order to maintain the effectiveness of cooperation and increase the efficiency of research. By timely and qualitatively combining the acquired knowledge, breakthrough and super-effective methods, technologies, solutions that constantly help to subdue local and global problems of mankind, to cover needs can come out of the activity of organizations.

The coherence of actions is based on scientific communication, which specifies two main stages: the stage of internal communication and the stage of external communication. From a brief definition of external communication, it can be argued that it covers the relationship in the scientist-society system. In turn, the internal stage is decisive in the process of all scientific communication.

The need finds its solution - a system of the global search for scientific information, the very idea of which is to combine existing resources: publications, magazines, books, research platforms into one large space of information, which allows you to use it without a separate search for a single resource request.

Some examples of existing implementations of the idea:

Elsevier - supports its own platform ScienceDirect (uses: Scopus - bibliographic and abstract database and tool for tracking citations of articles published in scientific journals; RELX - a global provider of information analytics and decision-making tools for professionals). CiteSeerX - indexes the scientific literature and automatically calculates the citation index to quantify the significance of individual publications - open-source platform. Google Scholar - attaches particular importance to the number of citations.

One of the issues, that Decision-support platform for international scientific and technical cooperation addresses, is an issue of access to scientific works of national and commercial enterprises of the world, which they distribute in publicly available specialized publications. But its more important part, which is embedded in the name itself - decision-support - solves the issue of decision-support itself.

Referring to Platform programm architecture, it is designed on the principle of multilayer. (UI (user interface), Building recommendations for choosing a publication, Communication with the database, Linking UI and building recommendations).

Steps of the interaction of these layers could be expanded like this:

1. Enter the required information
2. Up to 10,000 articles that are in the database are uploaded.
3. Filter the garbage. After this step, there will be fewer articles to analyze.
4. We form a rating of authors. Here are two approaches: to find whether the author was repeated somewhere in other articles loaded during the search, and to find the number of mentions of the author in the database.
5. We form a rating of contributors: the number of sources which interacted with the article is calculated.
6. We form a rating of freshness. The older the article, the lower the rating. Summarize the values obtained from the last three steps, applying the coefficients for each criterion to form a final rating

The decision-support part has its own “recommendations-building” layer. The recommendations are based on the analysis of the desired publications according to certain criteria. The criteria are based on an analysis of the author’s popularity in the database and extradition, publication date and contributors. So, the algorithm of recommendations-building layer lies in steps 4-7. The final rating is visualised as list of articles, which are recommended by the Platform that way. Recommendations allow anybody to make the right decision, who has a hesitation in the question of making some specific decision, that needs the different results of researches, etc.

Conclusions: in the field of information support, which allows finding the information very clearly and without mishits in the wide area of information, there are a lot of solutions from leading IT companies, commercial libraries of scientific publications. But progress means making new or making own ways in the realisation of global issues. The presented solution is a ready-to-use system and it is friendly to common users or researchers. Realisation uses open-source technologies, that includes API with shared documentation and offers different ways to connect with the database of scientific publications. These aspects allowed to realise specific recommendation system. Only the opportunity of making it real (with the help of open-source technologies) pushes the growth of knowledge-production and opens the ways to anybody to make our life better.

References

1. Qt Framework - Qt Features / – The way to access the

source: <https://www.qt.io/product/features>

2. Python 3 and PyQt 5. Applications development. 2-nd edition / Prochorenok, Dronov
3. REST API Design Rulebook: Designing Consistent RESTful Web Service Interfaces 1st Edition / Mark Masse

Kuzminykh V., Polishchuk K.

INTERPRETER OF ADVANCED DATABASE MANAGEMENT COMMANDS

Transactions are used in various subject areas: in banking systems, in ticketing and booking systems, etc. The main advantage of their usage is the guarantee of data consistency. But the ways to achieve these guarantees often negatively affect software performance due to the large number of locks [1], which is unacceptable for today's dynamic systems. Therefore, finding a new more effective approach to ensure data consistency and integrity within the transactions usage is an important area of research.

The *purpose* is to develop a database management algorithm that will allow flexible definition of rules for solving problems of parallel access to resources when using transactions.

In order to achieve the purpose of the research it is necessary to complete the following *tasks*:

- 1) To analyze the existing ways of solving the concurrency effects.
- 2) To present new effective approach to solve such effects without locking.
- 3) To develop an interpreter of advanced database management commands based on the presented approach.

Running several transactions, which may access the same data in the database, concurrently may cause errors or inconsistencies. The greater concurrency, the greater number of problems with data consistency may appear.

Among such concurrency issues are: lost updates; "dirty" reads; non-repeatable reads; phantom reads [2].

Most modern database systems provide the mechanism to solve these issues: while initiating new transaction, the isolation level is specified. This level guarantees that the problem, associated with this level, will be solved (Picture 1).

Concurrency issues Isolation level	Lost updates	Dirty reads	Non-repeatable reads	Phantoms
Serializable	don't occur	don't occur	don't occur	don't occur
Repeatable Read	don't occur	don't occur	don't occur	may occur
Read Committed	may occur	don't occur	may occur	may occur
Read Uncommitted	may occur	may occur	may occur	may occur

Picture 1. Concurrency issues and isolation levels solving these issues

Each isolation level is stronger than the one below. No higher isolation level allows an action forbidden by a lower one, the standard permits to run a transaction at an isolation level stronger than that requested [3]. This cascaded problem solving does not allow to solve the problem of phantom reads without solving all the other problems.

To ensure the guarantee of isolation, the highest isolation level should be used. For this purpose a database management system usually acquires locks on data which may result in a loss of concurrency. Depending on the access operation type and on the lock type, acquiring the lock may be blocked and postponed, if another transaction is holding a lock for that object. These locks are causing performance issues: waiting for resources to be released and deadlocks [4].

Even in cases of multi version concurrency control usage the problem of phantom reads cannot be solved separately from other problems. This will consume computer resources and block data in database. The latest research about making snapshot isolation serializable doesn't solve this issue either [5].

The new presented approach is to allow developer to specify which exactly concurrency issue needs to be solved within the current transaction (without the need of cascaded solving all other issues).

The ways of solving concurrency issues within the presented approach:

- Lost updates: allow rewrite or rollback transaction on detected issue.
- Non-repeatable read: allow non-repeatable read or use database snapshot.
- Phantom reads: allow reading phantoms or rollback on detected phantoms.

Detecting phantom reads is performed as follows: when executing the intermediate representation of command, for example, polish notation, save the description for the results for all "where" clauses in the transaction (number, versions and keys of all the returned records) and while committing the transaction execute polish notation for all "where" clauses again and compare the descriptions: if they are different, the phantom data appeared.

Moreover, the suggested approach allows to specify not only which separate concurrent issue to solve, but also the table for which this constraint will be applied. For example, the database management system will insure that within the current transaction the issue of lost updates will

be allowed for Table A, but not allowed for Table B (Picture 2). As well as, the rule not allowing phantom reads from Table C does not imply solving lost updates and non-repeatable reads from this table.

```
begin transaction with rules
    on lost update rewrite in A,
    on lost update rollback in B,
    on phantom data rollback in C,
    . . .
```

Picture 2. Transaction syntax due to the presented approach

Conclusions. The presented approach allows to specify rules for solving solving concurrency issues in a very flexible way: a developer may decide which exactly concurrency issues should be solved and for which tables. The issues are solved without using the cascading approach and locking, which minimises the resources usage for solving issues, makes data available even if used by the other transaction and mitigates the possibility of deadlocks.

References

1. Joseph M. Hellerstein, Michael Stonebraker, and James Hamilton: Architecture of a Database System, Foundations and Trends in Databases, volume 1. - NY.: Nov 2007. - p 141–259.
2. Abraham Silberschatz. Database system concepts, 6th edition. - NY.: McGraw-Hill, 2011. - 1376 P.
3. Philip A. Bernstein, Vassos Hadzilacos, and Nathan Goodman: Concurrency Control and Recovery in Database Systems. L.: Addison-Wesley, 1987. - 650 P.
4. Michael Melanson. Transactions: The Limits of Isolation. L.: Addison-Wesley, 2014.
5. Alan Fekete, Dimitrios Liarokapis, Elizabeth O'Neil. Making Snapshot Isolation Serializable. ACM Transactions on Database Systems, volume 30. - NY.: 2005. - p 492–528.

EXPERT SYSTEM FOR ASSESSING THE QUALITY OF INFORMATION

“Quality management is often called accurate information management, but this is often neglected” [1]. However, “information is not always the answer. Often information is only part of the problem” [2], so that we cannot trust information even from trusted sources [3, p. 3]. The problem of assessing the quality of information is explicitly or implicitly present in all information representations. It is both one of the complex, still fundamentally unresolved, and the most demanded tasks, the significance of which is still not appreciated at the proper level. The catastrophic state of this subject area is well reflected by enterprises [2, p. 76], 60% of enterprises suffer from poor data; and typically 10-30% of the data in organizational databases is not accurate; and industrial data errors can reach 75%; 70% of production orders are valued with poor quality data; 40% of the data in the credit risk management databases were incomplete; between 50% and 80% of abuse data are rated as inaccurate, incomplete, and ambiguous.

The close connection of information representations with the representation of the subject area implies uncertainty in many of its aspects. Obviously, it is difficult to focus on finding solutions to what is not defined. This creates the complexity and uncertainty of assessing the quality of information, which today must be considered in the "comprehensive analysis" $N = \text{"All"}$ [4] in the symbiosis of fundamental methods and technical capabilities. As a result, there is still low efficiency in the practical application of information technologies and systems. For example, even such “giants of opportunities” as Google, Yahoo still cannot provide services based on an assessment of the quality of information.

A brief overview of information quality research.

The state of representations of the domain of determining the quality of information is characterized by the presence of a large number of research papers presented by many publications. They consider the problems with a variety of previously isolated from each other various methodological approaches [2]. Representations of the quality of information inherit the problems of information representations, historically the structure of which has changed according to the semantic and value load. Therefore, we can distinguish three main categories of formed representations:

Consideration of the subject area as finding criteria for comparing different categories and types of information.

Assessment of the semantic and value load of information.

The tasks of determining the quality of information that are fundamentally defined at the junction of information and modern advanced fundamental scientific concepts and applied approaches - Big Data, Mashine Learning, etc.

The relevance reflects well the dynamics of more than 40 author's research thesis works on this topic: Kristo Ivanov (1972), Aksenov VV (1983), Afonichkin A.I. (1984), Craig W. Fisher (1999), Markus Helfert (2002), M. Pamela Neely (2002), IRIT ASKIRA GELMAN (2003), Hongjiang Xu (2003), Helinä Melkas (2004), Monica Scannapieco (2004), Cinzia Cappiello (2005), Karolyn Kerr (2006), John P. Slone (2006), Kapustin A.A. (2006), Dineev A.N. (2006), John Martin O'Donoghue (2007), Frau Dipl.-Ing. (2007), Shirlee-ann Knight BA (2007), Volkswirt Veit Koppen (2008), Paolo Missier (2008), Donald Ballou and Harald Pazer (2008), Gregory Hill (2009), Mouzhi Ge (2009), Greg Holland (2010), Olukayode Isaac Osesina (2010), Neal Gibson (2010), I. Rylsky (2010), Ahmed Abu Halimeh (2011), Tony O'Brien (2011), Ms Mary Levis (2011), Dr. Markus Helfert (2011), Mohamed A. Yakout (2012), Yinle Zhou (2012), Sokolova E.S. (2012), Shmarikhina E.S. (2012), Gonchar O.V. (2013), Soffi Westin (2014), Diana Hristova, MA (2015), Mohamad Gharib (2015). and etc.

Analysis of recent studies on which this work is based.

The need to compare different types and categories of information in determining the quality of information still presents a difficult, ambiguous task - a unified research structure is needed [2, p. IX]. Therefore, we cannot focus on many published works, even the latter, returning to the foundations of the origin of the problem. In [5, 6], four basic approaches to determining quality, which can serve as the basis for the development of a unified approach: 1) excellence; 2) value; 3) compliance with the specification; 4) compliance or exceeding customer expectations. Also, in view of the new “comprehensive” requirements $-N = \text{"All"}$ [4], increased requirements for an adequate informational representation of the subject area of knowledge, we apply the approach, [3] expanding the concept of information quality to the boundaries describing its indicators, as an integral component of their states. This approach has not yet revealed all the possibilities of its potential, but has already proved its effectiveness, for example, in the CoreZoid system [7], confirming the correctness of our chosen direction. As the basis for building an expert system, we considered the work of the famous scientist George Clear [8].

Non-decisive issues and the purpose of the work.

The limited use of highly specialized approximate methods for determining the quality of information, in connection with all-encompassing

requirements, does not allow for the effective solution of many practical problems. The uncertainty of many of the components of the problems of obtaining decisions in determining the quality of information can be resolved in the system of an adequate expert system, the creation of which requires the formation of its corresponding domain of knowledge.

Creation of an expert system for determining the quality of information.

If earlier we could limit ourselves to the choice of one or several approaches to create an expert system, today to get effective solutions it is necessary to look for a symbiosis of many approaches, taking into account additional factors:

- The search for solutions is considered by “coordination between two nodal points”. A V-shaped model is used as methods: formal approaches (analysis - decomposition of the subject area) and expert assessments (synthesis - solutions) are consistent, revealing a holistic approach.

- In representations of formal fundamental methods, the integrity problem has so far been ignored. Experts claimed that “integrity is unformalized”, however, it is permissible to transfer the concept of integrity to a new category of complexity by creating the following hierarchy of complexity: elements - systems - management - integrity. This is confirmed by a significant increase in work in the field of control theory [9]. Obtaining holistic ideas requires a review of axiomatics, resolving contradictions, taking into account the factor of human behavior, etc.

- In assessing the quality of information, we take into account six categories of the studied indicator X (X0, XI, XII, XT, XN, Xm), where X0 is the true value; XI is the statistical value; XII is the calculated value; XT - technological value; XN - regulatory value; Xm is consumer value.

- The model for determining the quality of information is based on the production principle of converting factors with additional conditions for taking into account the real world system and observers.

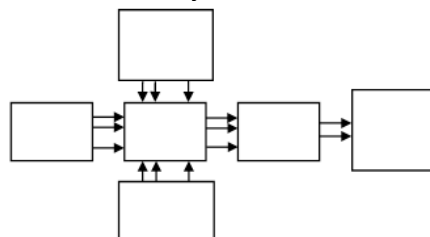


Figure 1. A holistic domain model

- The tasks of determining coefficients, optimizing and directly

determining the quality of information are solved by machine learning methods.

Findings.

One of the most important tasks in conditions of large, excessive volumes of information is to increase the efficiency of information technologies and systems in various applied tasks, which cannot be achieved without resolving the problems of determining the quality of information. In this situation, the right decision is to create expert systems, which are an intermediate stage in the revolutionary transition from information technology and systems to knowledge operating technologies. In turn, this implies the creation of integrated systems in which we consider reasonable, for which approaches to determining the quality of information are one of the factors to teach information technology and systems to "understand" information. As Corezoid’s experience shows, the introduction of an expert system for determining the quality of information can increase both efficiency,

References:

1. Исикава К. Японские методы управления качеством: Сокр. пер. с англ./Научн. ред. и авт. предисл. А. В. Гличев. - М.: Экономика, 1988. - 215 с.
2. Handbook of Data Quality. Ed. by S. Sadiq. Springer-Verlag Berlin Heidelberg Publ., 2013. 438 p. DOI:10.1007/978-3-642-36257-6.
3. Тараненко Р.А. Качество информации: обзор представлений в контексте целостной парадигмы / Руслан Анатольевич Тараненко. // УСиМ. – 2005. – №5. – С. 3–12.
4. Майер-Шенбергер В. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим/ Виктор Майер-Шенбергер, Кеннет Кукьер; пер. с англ. Инны Гайлюк. - М.: Манн, Иванов и Фербер, 2014. - 240 с.
5. Pipino L. L., Lee Y. W., Wang R. Y. Data quality assessment //Communication of the ACM.-2002.-Vol.45, № 4.-P.211-218.
6. Kahn B. K., Strong D. M., Wang R. Y. Information quality benchmarks: Product and service performance //Communication of the ACM.-2002.-Vol.45, № 4.-P.184-192.
7. Corezoid: и востали из пепла... гуманитарии. / URL: <http://bloggerator.org/page/corezoid-iz-pepla-gumanitarii-oblachnoe-programmirovanie>
8. Клир Дж. Системология. Автоматизация решения системных задач: Пер. с англ. - М.: Радио и связь, 1990. - 544 с.
9. Новиков Д.А. Кибернетика: Навигатор. История кибернетики, современное состояние, перспективы развития. - М.: ЛЕНАНД, 2016. - 160 с.

Kuzminykh V. Tymoshenko M.

AUTOMATION OF ASSESSMENT AND FORECASTING OF DEVELOPMENT OF HIGHER EDUCATIONAL INSTITUTIONS

In today's world, any organization needs to understand its place in the business environment. Now such information is obtained and analyzed exclusively by experts. As a solution to this problem, a system for assessing and forecasting the development of enterprises was developed. A successful area of application of such a system is the area of higher education institutions, due to the availability of open information about them, as well as the transparency of evaluation criteria.

Of course, it is necessary to analyze the activities of the university not only by economic parameters, for a complete picture it is necessary to identify other categories of parameters, namely scientific and quality parameters of education. Information for analysis can be found in open sources, namely on the websites of rating agencies. We managed to extract data collected by the largest rating systems, such as: QS World University Rankings, THE World University Rankings and Academic Ranking of World Universities, ARWU (Shanghai University Rankings). The Kaggle, the website for data science challenges, has data on these rankings for 2011-2016, and information for 2016-2019 can be found on the official websites of these agencies in the public domain.

Example of economic parameters: income, the ratio of the cost of equipment to the number of students.

Example of scientific parameters: publications, citations, influence abroad.

Example of quality parameters of education: an indicator of the ratio of the number of male students to female students, the number of students, the number of foreign students.

The system is divided into four main stages:

1. Data preparation
2. Comparison of the selected university with others by selected groups of parameters
3. Construction and training of a predictive model of university activities
4. The output of graphic information based on the results of the system

Data preparation

The first stage of data preparation is the cleaning of incorrect data. Thus, the main predictive variable is the final assessment of the university in the rankings, so all records with missing or incorrect data, such as zero

values, are deleted.

The second stage is the correction of incorrect data. Some less important variables may also contain erroneous data, but if you delete all the rows with them, the accuracy of the prediction model will drop significantly. Therefore, it is best to replace them with the average value for a given period.

The third and final stage is a data conversion. Some information about the organization may be stored as unsuitable for analysis. For example, the country in which the institution is located is of great importance for evaluating the activity of a higher education institution, but it is impossible to give the names of the countries to the model in text form, so they are numbered and broken down into other parameters as shown in Table 1. the predictive model will give a much better result. This type of transformation has almost no effect on different kinds of regressions, but it helps decision tree algorithms or other complex algorithms that use trees, like XGB or LGB to reach more precise.

	United States of America	United Kingdom	Denmark
Harvard University	1	0	0
University of Cambridge	0	1	0
Aalborg University	0	0	1
Princeton University	1	0	0

Table 1. Transformed names of countries of higher education institutions

Another type of conversion is to change unusual formats. For example, the rating of men to women is presented in the form of a percentage: "33:67", in which case you need to translate the data into the usual rating from 0 to 100, so the result of this example will be "33".

A common problem in data processing, which is present here, is the problem of multicollinearity between data. Unfortunately, we can't delete data to overcome it, because it will reduce the quality of the forecast.

Universities are compared with each other according to the collected parameters, their indicators are compiled into categories by which they can

be compared graphically or numerically.

Development forecasting

To predict the development of enterprises by parameters, three main types of methods were studied:

1. Extreme Gradient Boost, Light Gradient Boost
2. Multi-Layer Neural Network
3. Linear and LASSO regression

XGBoost and LGBost are machine learning algorithms that consistently take the lead in a large number of Kaggle competitions.

However, unfortunately for this task, they did not fit, MSE of their forecast was more than 10, which is a big error for this data. It is assumed that this happens due to a large number of parameters with not so many records. These algorithms actively use decision trees in their work, which give poor results under such conditions.

The multilayer neural network performed better, MSE reached 3.1185 when used, which is a pretty good result.

But using linear regression had to achieve MSE 2.1281, thanks to well-prepared data, but later a study was conducted, during which it became clear that the deterioration of raw data, also greatly reduces the efficiency of this algorithm, so it was decided to use LASSO regression.

Formula 1. LASSO regression

It is traditionally used when it comes to solving the problem of multicollinearity, and it is less sensitive to data errors, so it was possible to achieve an MSE of 1.0937.

Thus, automated analysis of the activities of higher education institutions is possible. The main categories of parameters by which they should be evaluated are economic, scientific, and quality parameters of education, which allows them to compare universities with each other, as well as to predict their development in these categories in a short time.

References:

1. Реалізація комплексного підходу до оцінки конкурентоспроможності ВНЗ і його продуктів 2013 / Гісіна Ольга Григорівна
2. Brown, M. G., Svenson, R. A. Measuring R&D productivity // Research technology management. — 1998. Vol. 41. — № 6. — P. 30.
3. Wilks D. S. Cluster analysis // International geophysics. — Academic press, 2011. — T. 100. — C. 603-616.

Mykhailova I., Tyschenko A.

WEB REGISTRY OF ELECTRONIC INFORMATION RESOURCES

Information resource (IR) – systematized information or knowledge that has value in a particular subject area and can be used by a person in their activities to achieve a certain goal [1]. These types of IRs, which are used by organizations or government entities, are stored in registries that allow them to work conveniently. The registry is an information and telecommunication system designed for registration, accounting, accumulation, processing and storage of information on the composition, content, location, conditions of access to electronic information resources and meeting the needs of legal entities and individuals in information services.

At the state level, there are different registries in which information with different types of access to it is stored. Similarly, organizations and individuals use registries to perform day-to-day tasks. Users can use currently many different out-of-the-box options. These are, for example, FossDoc, FlyDoc, Directum, Optime-WorkFlow, etc. However, most of these systems are proprietary, require paid customization, and have a surplus of features to meet simple needs. Thus, the issue of developing an easy-to-use web-registry with basic functions available free of charge to the general public is relevant.

To implement this task, the functionality of the software was determined:

- upload IR into the database;
- storage of meta-information on the downloaded IR (name, extension, size, tags, file description, date of upload);
- removal of IR from the database;
- removal of all IRs from the database;
- IR search based on document name, extension, tags, upload date;
- display of IR data to the user;
- display a list of IRs loaded in the database;
- download the chosen IR.

The multimodel database management system InterSystems Caché was used to implement this task. It allows a programmer to use object-oriented, relational, document-oriented and hierarchical models to work with data using built-in language to write business logic [2]. Business logic on the client side was written using TypeScript language [3], to display the user interface – Angular 9 [4] was used.

This system is a client-server software, which means that to use it

you only need Internet access and a browser. The architecture is shown in fig. 1.

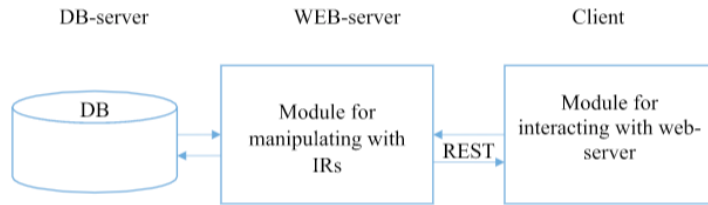


Figure 1 – Architecture of web-registry

Web-registry consists of the following modules:

- server module for manipulating IRs;
- client module for interacting with web-server;
- database of IRs.

The document manipulation module is a RESTful service that is written on ObjectScript language integrated into Caché database and is designed to save, return and delete documents. It interacts with the database. The input for this module are commands (GET, POST, DELETE, PUT) and IRs' text and metadata.

The interacting with web-server module is used to form commands and send/receive IRs and their meta-data to web-server.

The developed software allows the user to view the list of uploaded files and add new documents on the main page of the resource (fig. 2).

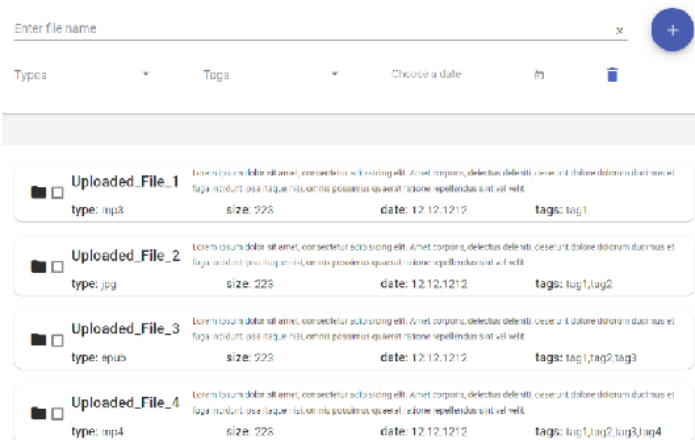


Figure 2 – Home page with a list of uploaded files

To upload a new document, the user must click the "+" button, then a window will appear (fig. 3), where the user must choose a file, specify tags and file description (optional), and click "Ok". After uploading the document, the window will close and the list of files will be updated. If the user decides not to upload the file, then he needs to close the window by clicking the "Cancel" button.

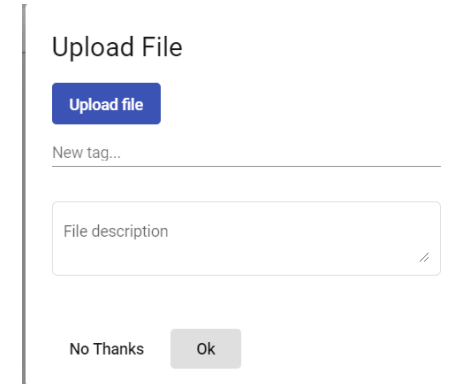


Figure 3 – Window for uploading a new file

To search by tags, upload date, file extension or IR's name, the user needs to enter keywords or values in the form fields on the main screen (fig. 2) and the list is automatically filtered depending on the search parameters entered by the user (fig. 4).

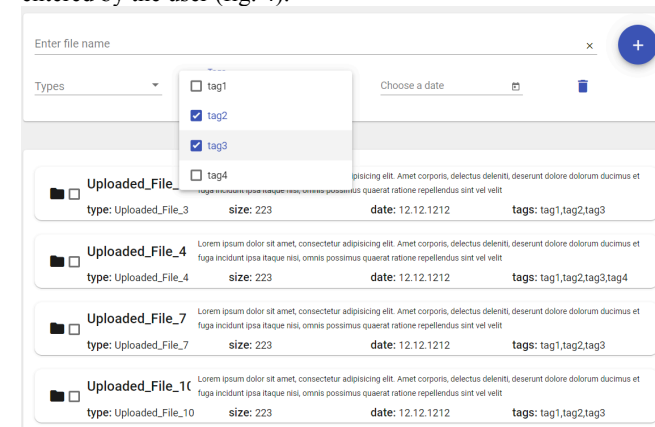


Figure 4 – List of files filtered by tags

Thus, a web application was created that allows the user to upload, search, download and delete information resources from the database. The advantage of this software is that the user has a single universal repository, to work with which he does not need any additional knowledge about the distribution of data depending on its format, all that is required of him is to enter the address of the web-resource and upload, find or download the existing IR.

References:

1. Про схвалення Стратегії розвитку інформаційного суспільства в Україні від 15 травня 2013 р. № 386-р [Electronic resource]. - Access mode: <https://zakon.rada.gov.ua/laws/show/386-2013-%D1%80/ed20130515#n21>
2. The Multi-Model Database: Enabling Business Transformational Applications [Electronic resource]. - Access mode: <https://www.intersystems.com/resources/detail/multi-model-database-enabling-business-transformational-applications/>
3. Typescript [Electronic resource]. - Access mode: <https://www.typescriptlang.org/>
4. Angular [Electronic resource]. - Access mode: <https://angular.io/docs>

CLOUD DATA STORAGE SERVICES

Currently, one of the key areas of information technology development is cloud computing. The cloud database management systems (DBMS) rapid growth in the popularity has been made possible by two coinciding factors. Databases were historically built using the only available at the time system of vertical scaling (scale-up), which required using increasingly powerful servers. In the early 2000s, the pinnacle of development in this direction were several powerful databases (Oracle, DB2, etc.) that operated on multiprocessor Unix-servers. But with the development of thin servers (1-2 U height) and especially blade servers in the same years, vertical scaling began to give way to cheaper horizontal scaling (scale-out), which increases server capacity by clustering relatively low-power servers of standard x86 architecture. As a result, the databases were "moved" to clustered configurations compiled from standard architecture servers or to hyperconvergent integrated systems.

Five years later, cloud technology entered the market. It became a significant part of the computer services market by 2010. And it turned out that the DBMS versions, originally designed for clusters, were prepared in advance for porting to infrastructures provided by the cloud services [1]. Multiplied by the unlimited scaling potential of server resources and storage capacity provided by cloud technologies, these two factors resulted into a synergistic effect that became the source of the cloud databases success. The first to migrate to the cloud storage were relational databases, soon after overcoming some technical difficulties, they were followed by non-relational.

Currently, all leading DBMS developers have a cloud version of their system. If you look at the latest Forrester Wave, dedicated to the cloud database market, among the leaders and their closest competitors you will find Oracle, SAP, IBM, Microsoft. Purely cloud vendors compete with these companies with a certain degree of success. There is a place on the cloud storage market for free DBMS. For example - MongoDB, which was even able to become one of the leaders according to Gartner Redis. (although, of course, them being "free" is even more conditional than placing your data in a corporate data center) [2],

Thus, the volume of the global DBMS market has increased due to cloud projects.

According to experts, the main contribution (about 68%) to this rise was made by cloud technologies. The most prominent of the manufacturers

were Microsoft and Amazon Web Services (AWS) which accounted for 75.5% of the market growth. The study showed a trend of DBMS transition to the cloud environment. It is expected that by 2022, approximately 75% of systems will be deployed or transferred to a cloud platform, and only 5% of solutions will be realized in the opposite direction - from the cloud to the local infrastructure.

Experts attribute this trend to the growing popularity of the SaaS business model (software as a service), as well as the practical value of using DBMS for analytical tasks. Customers implement new programs and transfer existing resources to the cloud services at an exponentially increasing rate. Companies are increasingly using cloud services for tasks related to analytics, machine learning, artificial intelligence, data lake and data storage.

According to analysts, in some cases, cloud services may be more expensive than traditional software, but the former provide such undeniable advantages as: eliminating the need for infrastructure management and the ability to more easily scale processes, as well as the cloud services often being a safer option.

Although in 2018 the costs for local DBMSs of Oracle, Microsoft, IBM and other vendors increased, the positive dynamics was achieved not due to new implementations, but due to forced software upgrades. Companies are forced to install new versions and additions to products in order to minimize risks. Most new DBMS projects run on the cloud technology. Under these conditions, infrastructure providers for cloud services and related services, which are based on it, become a new platform for data management.

Gartner claims that virtually all innovations in the database management system market are emerging within the cloud technology. Therefore, anyone who wants to take advantage of the latest developments should be present in this environment.

Cloud services providers form ecosystems around themselves, offering not only the services themselves, but also opportunities to interact with them and with competing products.

In the category of local databases the approach is different: here some products often perform several tasks, but rarely offer built-in mechanisms for integration with other products in the local environment.

On the fig. 1 you can see DBMS market leaders as per the end of the second quarter of 2019

Two DBMS types

The novelty of cloud databases (CDBMS), as well as other cloud software, lies in forgoing of the traditional "product" form of distribution.

The user is relieved of the need to invest in the purchase of the necessary hardware and software, instead gaining access to similar resources located in the data centers owned by the cloud services provider. Currently, two main approaches to the implementation of CDBMS are being utilized.

First, one can use the "Infrastructure as a service" (IaaS) approach and rent the necessary infrastructure resources. The cloud platform allows the client to reserve the required number of copies of the virtual machine for the time desired by the user, which supports the work of the DBMS allocated to the machines. Thus the client is able to set up the DBMS chosen by them on the rented virtual machines.

In the second case pre-prepared machine images are used. They are usually provided by DBMS manufacturers in conjunction with a cloud provider, such as Oracle Database 11g EE on Amazon EC.



CDBMS market leaders, 2-nd quarter of 2019 [3]

The size of the circle on the chart signifies the commercial success of the vendor's products in this area.

The second approach to CDBMS is based on the "Software as a service" (SaaS), where the DBMS is used as software, so it is given its own name "Database as a Service" (DBaaS). By providing DBaaS services, providers relieve customers of the need for any independent action with

instances of virtual machines. These functions are assumed by the provider, as well as all the responsibility for the installation and maintenance of the database. In this case, the user only has to make use of the ready-made solution and pay for its work. An example of such services is Amazon Web Services that offers an interface with the MySQL database.

The first option has well-known advantages and disadvantages of cloud solutions. The advantages include economic efficiency, unlimited (at least in theory) scalability, the absence of any problems with updating of the necessary equipment when new requirements arise, great flexibility and reliability. Disadvantages are that the customer cannot control data security limited by the service level agreement, problems with data networks can lead to downtime, data placement and software in the provider's data center causes attachment to it. When working with the DBaaS model, there is no need to buy software licenses, have a staff of specialists maintaining the database, the provider is responsible for the reliability of the database - to the extent that it is specified in the service level agreement. The main disadvantage of DBaaS is an even lower degree of data control.

When choosing one of the two options one should be governed by the scale of the task the company is facing. For small and medium-sized businesses in modern circumstances DBaaS would be the most optimal. Meanwhile, for a larger business it may be enough to rent IaaS equipment, but for a few very large companies working with huge amounts of data, it is still wiser to place databases within their own premises using their own equipment.

And, of course, much depends on the tasks to be solved. If your tasks are average you can use the assessment of Gartner analysts on the state of affairs of the operational DBMS (ODBMS) market. These are DBMSs designed to address a wide range of corporate transactional applications (ERP, CRM, etc.).

Cloud technology leaders. Leading DBMS providers are leaning towards the classic cloud model. For example, Amazon Relational Database (RDS) offers MySQL, Oracle and Microsoft SQL Server services. The simpler Amazon SimpleDB service is designed for tasks with low workload and complexity. Proponents of NoSQL can take advantage of Amazon DynamoDB, which works particularly well on solid state drives, and its replication process takes place in at least three physically spaced data centers. Notably, this service has the fastest growing demand in Amazon's history [4].

Oracle Database 12c is a part of the Cloud Application Foundation, which integrates the Oracle WebLogic application server, Coherence DBMS, and Oracle Enterprise Manager. Analysts from IDC even called this solution a qualitatively new integration trend that provides unique ways for

the application server to interact with the database, but how it differs from the integration approaches from ten or twenty years ago is still not very clear for the average consumer.

SAP is also actively developing in the field of cloud technologies: in addition to the HANA cloud platform, which primarily provides access to business applications, the DbaaS model through AWS is available in the Sybase database. SAP cofounder Jim Snape even claims that SAP is the only company in the world to offer a full line of cloud-based cloud systems. The experts call HANA the closest product to the Microsoft SQL Server 2014 with a built-in OLTP-engine "in memory", which is available in beta. Its backup functions are well integrated with Windows Azure, there is also a technology for creating "replicas" of data AlwaysOn Availability Groups.

DbaaS leaders. 12\$ million was quickly raised for DbaaS Cloudant investments from companies such as Rackspace (cloud version of Cloudant products), Samsung, Avalon Ventures, Toba Capital, the CIA's innovation division In-Q-Tel, etc. One of the directors of Devonshire Investors said that today there is a greater than ever market interest in managed database hosting, and mobile and web systems are today moving to the scalable model NoSQL-DBMS.

Xeround has found a good niche by developing a system for managing and deploying easily scalable MySQL databases on various cloud platforms and different technologies (AWS, CitrixRackspace, Joyent, HP, OpenStack). The most popular one today is NoSQL MongoDB and is available from numerous cloud providers, from AWS to Joyent. Garantia allows you to run NoSQL services Redis and Memcached in the AWS clouds. And the company StormDB refused to use the virtual infrastructure - its eponymous database is provided as a service directly from the physical servers of its own mini data center [4].

The original path has also been taken by the startup GenieDB, which released the product GenieDB Globally Distributed MySQL-as-a-Service (Linux-extension for MySQL 5.1-5.6). Its distinctive feature is the global synchronization of databases between geographically dispersed data centers and various cloud providers. At the same time, clients continue to work with their own version of MySQL. The authors of the technology have revealed they were inspired by the Google Spanner system, which uses GPS sensors to monitor the physical location of information! And another implementation of this idea is the TransLattice system, which differs from GenieDB only in the use of PostgreSQL database.

Conclusion. Data warehousing industry of the world is in its infancy, which gives hope for prioritizing the promotion of this technology both in scientific terms and in terms of commercialization of intellectual property. This is facilitated by the readiness of the free tools market for developing

web applications and web services in the programming languages like Java, Python or Ruby on Rails, widely used in cloud technologies, and there is an existing extensive practice of using these tools to develop web technologies for clouds.

References

1. Леонид Черняк СУБД уходят в облака. Как это получилось, и какое будущее их ждет?, Мск: https://www.cnews.ru/articles/2019-09-18_bazu_dannyh_uhodyat_v_oblakakak_eto (дата звернения: 13.03.2020).
2. Топ-10 стратегических технологий Gartner на 2018 год <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/> (дата звернения: 13.03.2020).
3. Forrester Research сообщает о финансовых результатах за третий квартал 2019 года <https://go.forrester.com/research/predictions/> (дата звернения: 13.03.2020).
4. Бобровский С.СУБД: облачные и “как сервис” PC Week/RE №30-31 (850-851) <https://www.itweek.ru/infrastructure/article/detail.php?ID=154194> (дата звернения: 13.03.2020).

Shushura O., Havrylko Y.

CONCEPTUAL MODELING OF FUZZY CONTROL

Approaches to the software development require using the conceptual modelling at the moment of task setting and design. The methodology of fuzzy control is widely used for the software development of complex systems control [1,2]. The development of conceptual models of subject areas in solving fuzzy control tasks requires special instruments that can formalize the ordinary and linguistic variables of control object, presenting constraints of the management.

Modern methodologies of conceptual modelling of subject areas, such as IDEF, DFD, ER and UML are most widely used, but they do not reflect fuzzy data [3-5]. Frames and semantic networks are widely used to describe product rules, but they do not allow to simulate fuzzy characteristics [6].

Fuzzy ER Model, Fuzzy IFO Model, Fuzzy EER Model and Fuzzy UML Data are proposed to adopt for fuzzy and imperfect information [7,8]. The considered methodologies are mainly intended for fuzzy data modelling and cannot describe the task of control.

Existing approaches to the modeling of fuzzy control tasks are usually represented by a rectangle, and arrows showing the input and output data [9]. Such approaches are too simple and do not allow to present the information about the variables relations.

The goal of the paper is to develop approach for the conceptual modelling which will provide description of fuzzy control in the graphic form.

In the paper suggested using the models:

- a conceptual diagram of the task of fuzzy control;
- a diagram of linguistic variables;
- a diagram of structure of fuzzy rules.

The conceptual diagram of the fuzzy control task is presented as a diagram that contains a graphical description of the ordinary input and output control, the related linguistic input and output variables, and the limitations. An example of a conceptual diagram is shown in Figure 1.

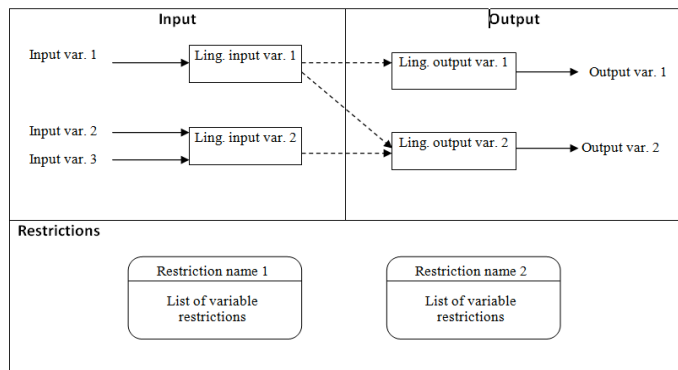


Figure 1 – An example of a conceptual diagram of fuzzy control

As we can see in Figure 1, the area of the diagram is divided into 3 parts: the input and output of the control and the restrictions. In the input and output areas, the names of the ordinary input and output variables are shown. The rectangles contain the linguistic variables and the arrows display the relationship ordinary variables with related linguistic variables. Dashed lines display the generalized inspiration of input linguistic variables for the values formation of the output linguistic variables, which need to be used when the rules of knowledge base is formed. In the third part of the diagram restrictions are presented in the form of rectangles with rounded corners.

Each linguistic variable must be detailed described as a frame in its graphical diagram. The generalized form of this diagram is shown in Figure 2.

Linguistic variable name	
Type	Input/ Output
Block of variables	
Name of usual variable	A set of valid values
Block of terms	
Term name	Characteristic of affiliation function
Description of new terms forming procedure	
Description of term affiliation forming procedure	

Figure 2 – Diagram of a linguistic variable

An example of relations between the linguistic variables is presented in Figure 3.

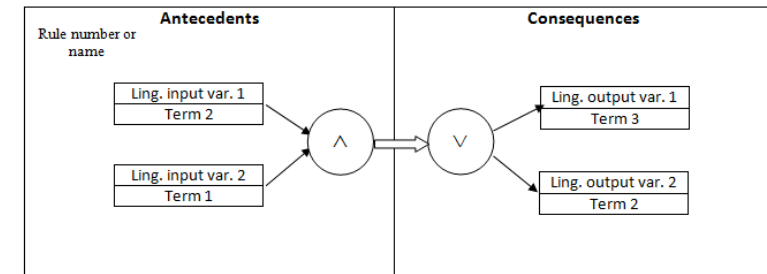


Figure 3 – An example of structure diagram for a rule of fuzzy inference

From Figure 3 it is clear that the structure diagram for a rule of fuzzy inference is divided into 2 parts: antecedents and consequences. In the antecedent's part, the linguistic variables and their terms are shown as rectangles. The antecedents are united into complex structure by logic operations of conjunction and disjunction, which are shown in the circles. The presentation of the structure of the rule inference is based on analogous.

Thus, approach to the conceptual modelling of the problem of fuzzy control by graphic schemes is presented in the paper. The approach includes a conceptual diagram of the control task, a diagram of linguistic variables, and a diagram of the structure of the rules of fuzzy inference. These diagrams allow to present information about the task of fuzzy control in terms of the subject area, which provides improved interaction between the customer and the software developer. The results of the paper may be used during the development of software for automation tasks based on fuzzy logic.

References

1. Zadeh L.A. Fuzzy sets. // Information and Control. – 1965. – Vol. 8. – PP. 338–353.
2. Шушурa О.М. Формалізація задачі управління в інтелектуальних інформаційних технологіях на принципах нечіткої логіки. / Шушурa О.М., Сторчак К.П., Бондарчук А.П., Золотухіна О.А. // Зв'язок, 2019. № 3. С. 3–7.
3. Г. Буч, Д. Рамбо, И. Якобсон. Краткая история UML // Язык UML. Руководство пользователя = The Unified Modeling Language User Guide. – 2-е. – М.: ДМК Пресс, 2006. – С. 14. – 496 с.
4. Лямец В.И. Основы общей теории систем и системный анализ

/ В.И. Лямец, В. И. Успенко. – Харьков : БУРУН и К ; Киев : КНТ, 2015. – 304 с.

5.Маклаков С.В. Создание информационных систем с AllFusion Modelling Suite / Маклаков С.В. –М.: Диалог-МИФИ, 2003. – 432с.

6.Дубовик С.Е. Использование семантических сетей, расширенных деревьями И/ИЛИ для представления структурно-динамических знаний в интеллектуальных системах / С.Е. Дубовик, Д.В. Сошников //Электронный журнал Труды МАИ. – 2003.– № 13.

7.Z.M. Ma A Literature Overview of Fuzzy Conceptual Data Modeling / Z.M. Ma, Li Yan // Journal of information science and engineering. – 2010. – № 26. – С. 427 – 441.

8.Peter Pin-Shan Chen The Entity-Relationship Model – Toward a Unified View of Data / Peter Pin-Shan Chen // ACM Transactions on Database Systems (TODS) : Сб. – Нью-Йорк: ACM, 1976. – Vol. 1. – P. 9-36.

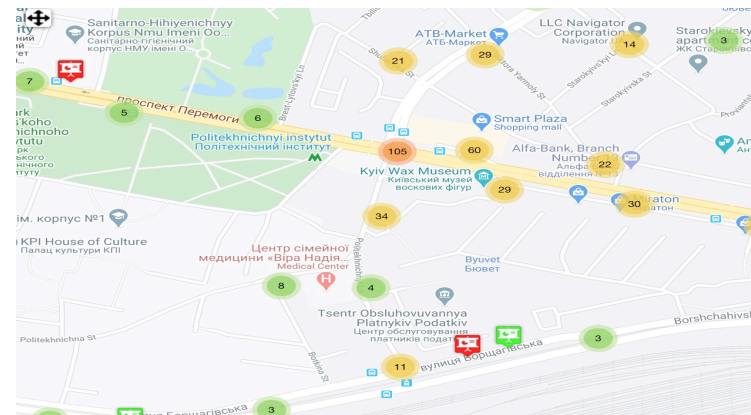
9.Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH. – СПб.: БХВ- Петербург, 2005. – 736 с.

Shaldenko O., Hulak O.

ANALYSIS OF MONITORING METHODS AND IDENTIFICATION OF ILLEGAL OUTDOOR ADVERTISING

Nowadays, outdoor advertising is very popular and highly effective. It combines many advantages: a wide audience, affordable cost and flexibility. Unlike advertising in print media, on radio and television, no one purposefully buys it, does not turn on the radio and television. It does not accompany entertainment programs to engage the audience. It is more often placed in places where the flow of people is quite high, because more people see the ad, the higher its effectiveness [1].

With the growing demand for outdoor advertising, the number of illegally placed ads has grown, which creates a new problem - pollution of the city (Picture 1) [4]. Utilities are actively fighting violators and involving inspectors from various organizations, such as KyivReklama and Blagoustriy in combating illegally placed advertisements.



Picture 1 - Map with marked illegally placed advertising

Since the beginning of 2017, a new advertising reform has been introduced in Kyiv, which describes the rules for placing ads, namely the ban on advertising in parking lots, road junctions, intersections, residential complexes, underground and overground crossings [2]. As a result, the amount of advertising has decreased 3 times. In order to maintain the level of income, it was decided to increase the cost of advertising, which

provoked the flow of illegal advertising.

Nowadays, due to the increase in the availability of mobile Internet, there is a tendency to use mobile applications to communicate with contact centers. At the moment, in order to report unauthorized advertising, you need to call the dispatcher of the Advertising Department of the Kyiv City State Administration, while the application allows user to avoid the need to make a phone call, which can result in long waits, unavailability, waste money. In addition, the user is not always in silence and has the opportunity to talk on the phone, which makes using the mobile application more convenient to make a call. Kyiv is a tourist city, so in order to improve its appearance, first of all it is necessary to get rid of illegally placed signs that cover historic buildings and damage the facade of buildings. Today, such advertising is disgusting not only to tourists but also to the residents of the city, and therefore people are becoming more aware, the number of those who want to improve the appearance of the city is growing. This is due to the relevance of the topic of application development for monitoring and identification of illegal outdoor advertising.

The goal is to create a mobile application for monitoring and identification of illegally placed ads, the implementation of which will allow inspectors of utilities, as well as residents of Kiev to clean their city from illegal advertising.

The fight against advertising is a problem not only in Ukraine but also in many other countries. For example, in June 2013, Toronto Mayor Rob Ford personally disrupted illegal ads from bus stops as part of the Clean Toronto Together campaign. It was assumed that representatives of the municipality will patrol the main streets of the city all summer, clean the posts and boxes of utilities, bus stops and mailboxes. And at the same time to issue fines to violators in the amount of 300 to 500 dollars. At the same time, advertisements posted by the residents of the district with information about lost property, a garage sale or a religious event were not subject to a fine. In addition, Illegal Signs volunteers report illegal objects to the City Hall (in Toronto, in particular, large-scale advertising on outdoor media is prohibited), after which violating companies dismantle the object and simply move it to another part of the city. They have managed to place their employees in all major companies that offer advertising, and they report to Illegal Signs prompt and accurate information about new illegal designs. But to launch such a campaign requires a lot of money and people willing to help in their spare time.

In the spring of 2012, activists in the American city of Auckland Park began working with the local municipality in an attempt to clean the city from illegal advertising.

The methods of struggle in neighboring Hollywood were analyzed.

There was a competition among residents with a prize of \$500. The award was given to the one who collected the most illegal ads. The streets of the city were cleared, but after a while the ads began to appear again. Then the authorities decided to try the autodial method. Its essence is simple: the phone number specified in the advertisement, every 30 seconds, the robot calls, thereby blocking the line for others, and reads the recorded message. It said that the autodial would stop as soon as the fine was paid. As a result, such calls reduced the number of illegal ads by 70%, while the operation with telephone "terrorists" cost only \$300. The low cost of this method is a great advantage, but in Ukraine buying a phone number is a low-cost procedure, so the same ads will appear, but with different numbers and over time, volunteers get tired of reporting illegally placed numbers and constantly calling the dispatchers of the Office of Advertising.

A group of Illegal Billboards volunteers is operating in New York. Their goal is to determine which advertisements and structures in the city are illegal, and to get them removed from the streets. According to activists, half of the signs in New York are placed illegally. Outdoor advertising has become a de facto black market run by corporations, and they are not stopped by official fines. Therefore, the diligent work of activists is needed, who investigate all suspicious signs and seek their removal through official mechanisms. Illegal Billboards teaches everyone how to fight illegal advertising. To do this, they hold regular seminars and post step-by-step instructions on their website. The main secret is to know which laws to refer to, as well as what information to report to which authorities. Even in cases where city officials confirm the legality of an advertisement, activists check their data. The problem of New York is similar to the problem of Kyiv, because in crowded places illegal advertising is many times more than legal. But there are fewer volunteers willing to spend their free time. So, this method is far from the reality of Kyiv.

The problem is more similar in Greece, where, as in Kyiv, there is a problem not only with advertisements, but also with illegal constructions, such as our advertising trailers. The existing law in Greece, which prohibits advertising along roads, was complied with only during the 2004 Athens Olympics. Residents of the capital have long complained about the mayor, who was suspected of colluding with illegal advertisers [3]. As a result, in 2010 they managed to get the Ministry of Transport and Infrastructure to launch a website and a special application for the iPhone, through which you can report illegal advertising.

Based on this method, which is the most optimal for the people of our country, we developed our own solution. A person who wants to report illegal advertising must take a few minutes to take a picture of the illegal advertising, download it to the application, set the geolocation and provide

the phone numbers listed in the ad. Authorized companies then review, track, and eliminate all illegal ads. The user will receive push notifications about the change of the ad status.

The purpose of the software product is to create an application for the elimination of advertising, display a map with the offenders marked on it, monitoring the status of the created applications.

To ensure the interaction of those wishing to remove the illegally placed ad and the supervisor, who will check whether the phone number corresponds to the information provided in the ad, the system includes applications for activists and utilities for different platforms, system server and application for supervisors.

In this article, a solution to the problem of monitoring and identification of illegally placed outdoor advertising was proposed and described. As a result of the analysis of existing methods of solving the problem in different developed countries, our own algorithm was formed: to form an application for illegally placed outdoor advertising using the application, send it to the supervisor, the supervisor checks the application and blocks the phone number.

The solution has been successfully applied to the system for eliminating illegal advertisements and will soon be used by Kyiv utilities.

Bibliography

1. New rules for advertising in Kiev: everything you need to know [Electronic resource] – Resource access mode: <https://dengi.informator.ua/2018/08/09/novye-pravila-razmeshheniya-reklamy-v-kieve-vse-chto-nuzhno-znat/>.

2. What to do if advertising on the streets of Kiev interferes [Electronic resource] – Resource access mode: <https://kiev.informator.ua/2018/08/09/chto-delat-esli-meshaet-reklama-na-ulitsah-kieva/>

3. City to clamp down on illegal outdoor advertisers [Electronic resource] – Resource access mode: https://www.joburg.org.za/media_Newsroom/Pages/2016%20%20&%202015%20Articles/City-to-clamp-down-on-illegal-outdoor-advertisers.aspx.

4. Map with marked illegally placed advertising [Electronic resource] – Resource access mode: <http://db.kievreklama.kiev.ua/db/rz.dll>

Sydorenko I.U., Kryvda O.V., Horodetskyi M.
**PROCESSING OF LARGE ARRAYS OF DATA BY
 GAUS INTERPOLATION METHODS**

Abstract. The problem of interpolation is especially relevant today, as the development of science and technology needs the processing of large data sets.

Large research institutes, firms and companies are investing heavily in the development of this area. Well-known companies and corporations include Google, Yahoo! and other. If you need to reduce the processing time and increase the accuracy of the result, there is a need to create an apparatus for interpolation data processing by analysing the problems of existing classical methods and improving modern ones.

The problem of obtaining values of the interpolation function with minimal errors arises in a large number of areas of human activity, such as geometric modelling of objects and processes, development, construction of real scenes (3D images), aircraft, banking, medicine, and others.

Analysis of recent research. Much attention is paid to fast data processing nowadays. Since there is a problem with the application of classical interpolation methods in the processing of large data sets, it was decided to use the Gaussian function for this problem. In [1] the concept of Gaussian function was introduced, formulas were given and approaches to interpolation were described. The parameterization of the Gaussian function is considered in [2]. The Gaussian interpolation function as a means of mobile data analysis is shown in [3].

The purpose of research. The aim of the work is to create an apparatus of Gaussian interpolation function by changing the variable parameter, which reduced the calculation error when processing large data sets, and based on mathematical principles to create a software product that allows to analyze Gaussian methods and automatically estimate the interpolation error.

Presenting your own solution. Consider the Gaussian interpolation formula

$$[1]: \quad G(x) = \tilde{y}_1 e^{-\alpha(x-x_1)^2} + \tilde{y}_2 e^{-\alpha(x-x_2)^2} + \dots + \tilde{y}_n e^{-\alpha(x-x_n)^2},$$

where α is the variable coefficient. In the classical case, this coefficient is written:

$$\alpha = \frac{\pi(n-1)}{(x_{max} - x_{min})^2},$$

where x_{max} , x_{min} — maximum and minimum value of the argument x .

In previous studies, this type of coefficient α was used. In this work,

Estimation of algorithm error:

Lagrange 0,5891638014699470
 Gaussian Parametric Normal 0,0000103527652694

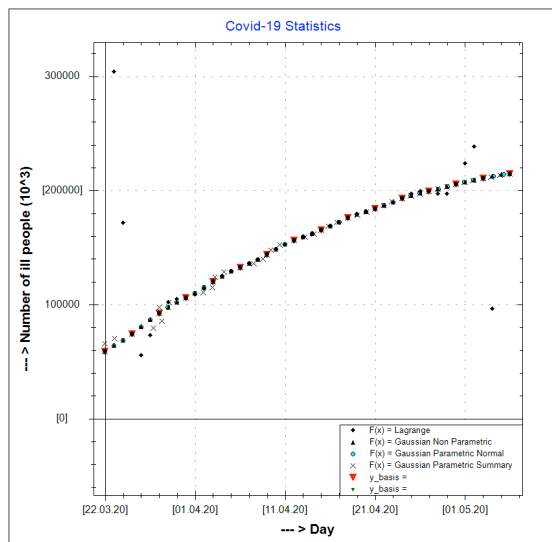


Figure 2 — Interpolation of the COVID-19 function in Italy for the period from 22.03.2020 to 07.05.2020

$X \in [0,1;10,1]$. Number of points = 15

Estimation of algorithm error:

Lagrange 0,1124218547921890
 Gaussian Parametric Normal 0,0000023435417150

Conclusions. Given a large number of interpolation nodes (more than 15) and a variable step, interpolation classical interpolation methods have large deviations from the actual result, and therefore it is necessary to use other methods, such as the Gaussian interpolation function.

The error of the Gaussian method can be reduced by influencing the variable coefficient α .

The Gaussian method proved to be good when working with arrays

of experimental data, which was confirmed by the example of interpolation of the spread of COVID-19 in Ukraine and Italy.

References

1. Badaev Y.I., Sidorenko J.V. Implementation of the Gaussian function interpolation method and comparative analysis. *Applied geometry and engineering graphics*. K.: KDTUBA, 1998, Issue. 63. P. 33-37.
2. Sidorenko Ju.V., Satskova A.V. Computer implementation of various methods of parameterization of Gaussian interpolation function. *Applied geometry and engineering graphics*. K.: KDTUBA, 2003. Issue 72. P. 63-67.
3. Gorodetsky M.V., Sidorenko Ju.V. Gaussian interpolation function as a means of mobile data analysis. Modern problems of scientific support of energy: Proceedings of the XVIII International scientific-practical conference of young scientists and students, Kyiv, April 23-26, 2020. In 2 vols. : "Igor Sikorsky Kyiv Polytechnic Institute", 2020. Vol.2. P. 88.

CONCEPT MAPS IN THE INDIVIDUAL LEARNING ENVIRONMENT

Typically, when learning new material, you need to write down the basic concepts and theses in order to further repeat the material, and for better memorization. However, simple summaries may be too voluminous, they may not be very convenient to search for material on a particular topic, and may not always be easy to understand the connection between different topics, so there is a problem of organizing the display of knowledge in a way that is convenient for learning new material, or repetition of the basic theses already studied. The display method should be informative, flexible in construction, show links between the facts, but not overloaded with data. For this work was explored the relative benefits of hierarchies and networks, it was noted that the benefits are largely complementary: hierarchies provide users with some understanding of central topics, allowing them to develop a better overview of information; while networks allow people to glean concrete information from the representation rather than needing to extensively read individual documents, mindmaps combine their properties and are perfectly suited to meet these goals. [1]

Mindmaps - diagrams that show the relationship between concepts (ideas), consists of nodes and lines with superscriptions between them. Such maps group facts and ideas around a central concept, consist of words and pictures, and visually organize complex information. They can be used to build ontologies, as well as to formalize knowledge. Maps - a graphical tool used to organize knowledge of a particular subject area, is a tree structure that reflects semantic or other connections between segments of information. The shape of the concept map is a radial tree, although in addition to hierarchical connections, additional edges of associations between different elements of the structure can also be used here. Due to the radial shape mind maps proved to be quite convenient for the presentation of information structures and became widespread as a means of visualizing ideas. [2] On such maps, each part (idea, concept) is connected to the other, with the help of such connections it is easy to understand its relation to the main idea, or its role in the system as a whole. Mindmaps help develop logical thinking and better understand information as they show how individual ideas are formed as a whole.


The main reason for the effectiveness of maps is that they mimic the way our brain organizes and stores information. We think of not numbered lists, but of ideas that are interconnected in different ways, which is why concept maps are better suited to learning than ordinary notes. Using this

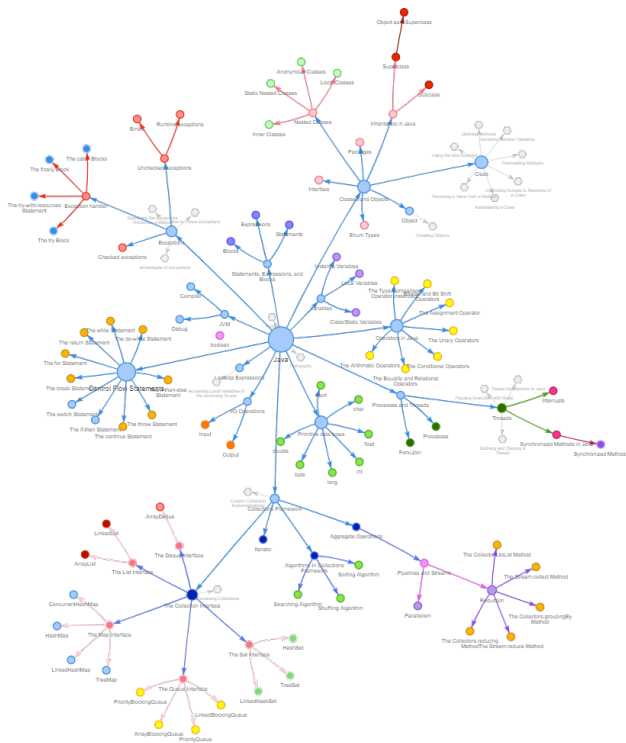
kind of notes, different parts of the brain are activated, parts that are responsible for creative and analytical thinking work together to improve material comprehension and memory. If you use different colors and nontrivial patterns when building maps, it helps to generate positive emotions and makes learning easier. Also the efficiency of the way mind maps display information is providing deep exploration of knowledges, this allows student to iteratively reduce confusion, slowly exploit every detail. In that way such hierarchical structures helps to navigate from more general data to specific one, which helps not to lose orientation in big amount of information.

Conceptual maps are used for:

- taking notes and summarizing key points
- assistance in creating a shared vision and understanding in sharing knowledge
- deepening the understanding of the individual elements of the system and the relationship between them
- when explaining complex ideas to experts
- software design
- structuring and visualization of knowledge
- assessment of students' understanding of the objects being studied

Usually when studying some material with the help of mindmaps the student has to build maps for himself, this helps to better understand the different parts of the subject area and their interaction. However, for beginners, building effective concept maps can be a difficult task, so it makes sense to have a methodology in which concept maps are built by a teacher who understands the subject area well and has experience with maps. These maps help you navigate the course material more quickly, and can also serve as an effective basis for repeating a course you have already completed. [3]

Good example of using concept map for displaying significant amount of information shown on  [4]



Img 1. – Concept map of Java language

A significant advantage of concept maps is the possibility of their semi-automatic generation. An expert in designing a training course can identify key concepts that will build the map, it will greatly simplify the understanding of the complex ideas outlined in the material, as well as simplify the memorization of the material, because the scheme with key theses is better than just text educational material. Building a concept map should pay attention not only to ideas, but also to establishing the correct connections between them, since they reflect the interaction of different elements of the system with each other.

Consequently, concept maps serve as a tool for learning and learning new knowledge, their range of applications is quite wide, and formal writing simplifies their construction, so they can be effectively used in the

learning process. Understanding the links between different ideas and concepts is a very important part of the learning process, so concept maps are better than ordinary abstracts for learning new material

References

1. Tytenko, SV Generation of test tasks in the system of distance learning based on the model of formalization of didactic text / SV Tytenko // Scientific news of NTUU "KPI". - 2009. - № 1 (63). - P. 47–57.
2. Tytenko, S. V. INTERACTIVE CONCEPT MAPS IN ONTOLOGY-ORIENTED INFORMATION AND LEARNING WEB-SYSTEMS. KPI Science News, no. 2, pp. 24–36, 2019. doi:10.20535/kpi-sn.2019.2.167515
3. Novak, J.D. (2009). Learning, Creating, and Using Knowledge: Concept Maps as Facilitative Tools in Schools and Corporations (2nd ed.).
4. Java: Semantic summary <http://semantic-portal.net/java/summary?full>

AUTOMATED CONSTRUCTION OF TEST TASKS BASED ON DECOMPOSITION OF EDUCATIONAL TEXT

Every day the amount of information for learning increases and the learning process increasingly becomes a distance form, which makes it problematic to quickly and efficiently assess the quality of acquired knowledge. The priority of the development of Ukrainian education is the creation of tools for distance learning and testing. One of the most common forms of distance testing is computer-based testing. The life cycle of test tasks in the educational Web-system includes such stages as preparation, submission and evaluation.

Pedagogical tests can perform such functions as diagnostic, educational, upbringing, developmental, motivational and control training. In order to ensure the creation of such tests, you can use different approaches, including automated construction of test tasks, but as a rule, the stage of preparation of test tasks remains poorly automated and requires direct work of the teacher [1].

The advantages of the automated system will be that this type of testing will reduce the time of checking the answers, will allow you to adjust the complexity depending on the correctness of the passage, get objective estimates excluding the human factor, the ability to automatically create statistical estimates of a particular group of students

Among the known means of automatic generation of test tasks should be noted the methods of parameterized tests and the use of semantic networks. The main disadvantages of these approaches are the creation of tasks with low pedagogical value, the lack of analysis of the complexity of the tasks. Therefore, there is a problem of developing software aimed at solving this problem.

The solution to the problem of constructing test tasks can be an ontological approach. This approach is often used to combine knowledge of subject areas into a single information structure. Since the data will be formalized on the basis of the concept-thesis model, which is developed taking into account the specific requirements of the educational process, the system will be able to auto-generate tests based on aspects. The level of information studied by the student will be divided into 3 categories

1. I remembered, I understood - low
2. Memorized, understood, perceived - average
3. Memorized, understood, perceived, analyzed - high

Certain tests on the topic will be from the newly studied topic, but

the rest will be repeated from the previous topics so that the student acquires more knowledge and has a holistic perception of the subject.

The concept-thesis model assumes an unambiguous connection between semantic data and the educational text. This provides an opportunity to manage the learning process and the process of control of knowledge and their adaptation.

In this case, we have information about the concepts and their location in the training material. The number of semantic units related to the concept is easily calculated for each educational fragment.

Tests are generated on the basis of semantic data of the conceptual and thesis knowledge base. The simplest scheme without taking into account additional characteristics (classification and level of importance of concepts and theses) looks like this:

- Random selection of the thesis, which will be the control of knowledge, from the outlined range of semantic data for a given area of educational material - control thesis;
- Definition of the concept to which the control thesis refers - the control concept;
- Random sampling, alternative theses from the whole field of theses of the given course (discipline or subject area), which do not concern the control concept.
- Construction of the problem according to the control concept, control thesis, alternative theses. [2]

The developed method must guarantee the following conditions:

1. Lexical intelligibility of test tasks
2. Ability to build tests of different types
3. Possibility of flexible testing depending on the purpose and strategy

Among the areas of further research is a detailed formalization of the solution of the problem of data adequacy for the formation of the test, improving the quality of test tasks and building a new task template. Particular attention should be paid to additional tools to verify the achievement of cognitive goals, according to the assessment scale

Thus, the automated construction of test tasks based on the concept-thesis model of formalization of the didactic text will provide an opportunity to generate test tasks of different complexity and analyze the quality of knowledge acquisition by students in different sections. With the development of flexible testing, it will be possible to choose a new testing strategy and goal for each section based on the test results.

References

1. Tytenko, SV Generation of test tasks in the system of distance

learning based on the model of formalization of didactic text / SV Tytenko // Scientific news of NTUU "KPI". - 2009. - № 1 (63). - P. 47–57.

2. Tytenko SV, Gagarin OO Practical implementation of testing automation technology based on conceptual and thesis model. Education and Virtuality - 2006. Proceedings of the 10th International Conference of the Ukrainian Association of Distance Education / Under Society. ed. V.A. Grebenyuk, Dr. Kinshuk, VV Sementsa.– Kharkiv-Yalta: UADO, 2006.– P. 401- 412.

Yerokhina A.

COMPARISON OF BINARY METHODS OF MULTI-LABEL CLASSIFICATION

The problem of automation has recently become relevant, because advantages over mechanical and routine work are significant: increased productivity, improved quality, time savings and increased profits. Particular progress has been made in the field of machine learning, which has been developing rapidly over the last 10 years and shows high results and is used in search engine generation, marketing information analysis, cloud server management, machine translation, image recognition of photos and videos, detection of spam at mailbox and social networks, plotting maps, and even identifying music that is playing alongside the user.

The purpose of this article is to study, analyze and compare binary methods of multi-label classification on the example of the problem of predicting genres of films based on descriptions of their plots.

The main technique for solving this problem is a method that determines a set of appropriate labels for a new instance of data. Tactics for predicting multiple labels differ from the classical approach of multiclass classification by the source data generated by the model. If the traditional classifier returns only one value, the multi-label must create a vector of initial values. This method of machine learning is used when each object is represented by one instance and is associated with several class labels at the same time.

The literature considers different approaches to solving this problem [1]. For example, to classify films by genre from the site "imdb.com", the developers used ensembles of binary classifiers, one of which was the popular method of Binary Relevance [2]. This method is an intuitive solution for learning on objects with many labels. It creates a number of independent binary classifiers separately for each label (Figure 1).

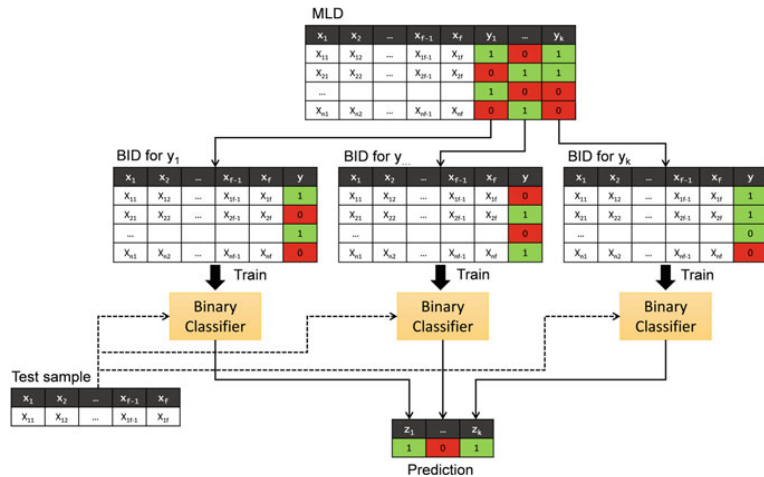


Figure 1 — Schematic of the Binary Relevance method [1]

Ignoring correlations between labels is a significant disadvantage, but among the advantages: ease of interpretation and use, the absence of hyperparameters.

Other researchers on the same data set preferred the Classifier Chains method [3], which combines the computational efficiency of the Binary Relevance method and can take into account the relationships between labels. The method constructs a binary classifier for each label, using the source information of the previous classifier as input for the current one. This algorithm usually gives high results compared to Binary Relevance. In the study, the developers compare Classifier Chains with a different approach to solving the problem of multi-label classification. Label Powerset is a simple method that treats each unique set of labels in the original training sample as one new class in the transformed data (Figure 2). Any multi-label classification algorithms can be applied to the new task.

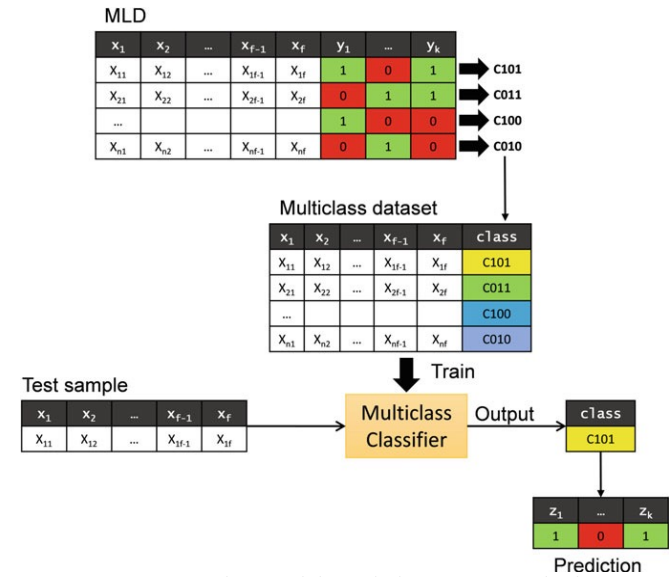


Figure 2 — Scheme of the Label Powerset method [1]

The class provided by the algorithm in the new problem uniquely corresponds to a certain set of labels in the original problem. One of the problems with the Label Powerset method is that after converting data, most of the new classes contain very few objects, and the distribution of objects in the new classes is extremely unbalanced.

This study analyzes the approaches of different authors to multi-label classification and compares the methods described by them to predict the genres of films on short text descriptions of plots. It should be noted that these approaches work as a "wrapper" over the basic model, so LogisticRegression was chosen. This is a statistical function used to predict the probability of an event by comparing it with the logistic curve. This regression gives the answer in the form of the probability of a binary event (1 or 0).

Comparisons were made among such classification approaches as Binary Relevance, Classifier Chains, Label Powerset and RAKelD. The last of these is a method that generates random subsets of labels by training the classical classifier on each subset. The advantages of this method are simplicity of calculations, more predictive capabilities, the ability to take into account the dependencies of labels. The disadvantage of this approach is the inability to use previously undefined labels for classification [4].

When working with models, the GridSearch method from the sklearn package was used to find the best hyperparameter C of the logistic regression model. This method sorts through the possible values of the parameter that needs to be optimized, and each time trains the model with new values. As a result, the option where the accuracy of the model was the highest is chosen. Variable C is a parameter that controls the effect of regularization (method of counteracting relearning) on the model. Below are graphs of the dependence of the metric F1, which is the average harmonic accuracy and completeness, on the hyperparameter C (Figure 3). It can be seen that on the last of the graphs at very small values of C the value of the metric F1 is in a wide range and the accuracy of the model is unstable.

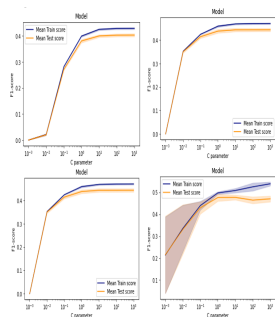


Figure 3 — Graphs of the dependence of the F1 metric on the parameter C for the methods Binary Relevance, Classifier Chains, Label Powerset and RAKelD

After selecting the optimized hyperparameter, the values of the main metrics for all multi-label classification methods were calculated. The results are shown in table 1.

Table 1 — Evaluation of model results

Metrics	Binary Relevance	Classifier Chains	Label Powerset	RAkelD
F1-micro	0.297	0.264	0.258	0.367
ROC	0.582	0.538	0.551	0.528
Hamming	0.257	0.210	0.202	0.126
Accuracy	0.059	0.080	0.125	0.226

It is easy to see that the RAKelD method gives the best results on the basic F1-micro metric. RAKelD generates random subsets of labels by teaching a multi-tag classifier on each subset. The Label Powerset method, unlike the previous one, encounters fewer combinations while retaining label correlation information.

During the study, the disadvantages and advantages of such binary

multi-tag classification methods as Binary Relevance, Classifier Chains, Label Powerset and RAKelD were analyzed and identified. As a result, it turned out that the RAKelD approach shows the best results on the problem of predicting genres by a short textual description of the film's plot.

References

1. Herrera F. Multilabel classification: problem analysis, metrics and techniques / F. Herrera, F. Charte, A. J. Rivera, M. J. del Jesus // Springer International Publishing Switzerland, 2016. — P. 65-78.
2. Godbole S. Discriminative methods formulti-labeled classification / S. Godbole, S. Sarawagi // Adv.Knowl. Discov. Data Mining, 2004. — P. 22-30.
3. Read J. Classifier chains for multi-label classification [Электронный ресурс] / J. Read, B. Pfahringer, G. Holmes, E. Frank // Department of Computer Science, 2011. — P. 5-16. — Режим доступа: https://www.researchgate.net/publication/227319967_Classifier_Chains_for_Multi-label_Classification
4. Tsoumakas G. Random k-Labelsets: an ensemble method for multilabel classification [Электронный ресурс] / G. Tsoumakas, I. Vlahavas. — 2007, vol. 4701. — P. 406-417. — Режим доступа: https://link.springer.com/chapter/10.1007/978-3-540-74958-5_38

Scientific publication

**MODERN ASPECTS OF SOFTWARE
DEVELOPMENT**

Proceedings of
VII Scientific and Practical Virtual Conference of
Software Development Specialists

June, 1, 2020