Implementation of analytical tools in the scientometric platform for the formation of consolidated information

**Oleksandr SHERSHUN, Iryna ZINCHENKO, Zhanna TYTURENKO, Anastasiia VOLKOVA ,
Olga OLSHEVSKA**

Odessa National University of Technology

shershunchik17@gmail.com, janettrnk@gmail.com, volkovanastasiia.w@gmail.com,
olshevska.olga@gmail.com

Abstract: The article analyzes the main problems of the scientometric area by the criteria for HEIs scientists carried out using the data mining methods. The main purpose of any research is to share achievements and accomplishments with society. However, each time a scientist publishes a he/she will be curious about how it's being received. Introduction of a category or criteria is a basic component of evaluating any activity, in particular scientific, and allows the unbiased consideration of the activities for building a balanced segmental circle. This approach was chosen to build the Odesa National University of Technology (ONUT) sciencetometric system, which allows researchers to have a quick and easy tool for identifying how much and what type of attention a research output has received. The software was developed in the PyCharm and DataGrip development environment. As a result of the research, a web application was created that fully meets the requirements of Odesa National University of Technology Scientific and Technical Library. The analysis of the subject area is carried out. The database was designed and developed using PostgreSQL DBMS version 10. Software solutions for the tasks were developed in the Atom environment using the Python programming language and the Django framework. The results are methods and functions that organize the operation of the system and the interaction of the application server with the database server.

Keywords: computing information technology, information management system, scientific methods, web, database.

1. Introduction

To assess the effectiveness of scientific activity, an important place belongs to scientometrics - a field of research that studies cognitive communications in science by the frequency of citations of scientific papers and their authors.

Previously, scientometric data were obtained and updated manually by the Coordinating Center of Scientific Journal Publishing department for each scientist of Odesa National University of Technology (ex. Odessa National Academy of Food Technologies), now this process has been automated. To structure the scientometric data, a web application was developed by IT and library specialists, which provides software support for the preservation of scientometric data of the teaching staff of ONUT.

Proper structuring of data helps the staff of the Coordinating Center of Scientific Journal Publishing of the Scientific and Technical Library of the Odesa National University of Technology to work with them more efficiently and quickly.

The S2M platform helps save a huge amount of time spent on routine tasks. The software also allows to generate reports on publishing activity by departments and institutes.

In the digital age, academic libraries and accountants cannot ignore the importance of multimedia products, which should be the most important means of communication. Accordingly, with the introduction of new technologies, the system of library management and document management is completely changing. Proper library management, taking into considering the activities of stakeholders, helps employees to better and more efficiently serve visitors and perform internal tasks such as reporting documents.

2. Sciencetometric method

Today, information technology in the scientometric field plays a very important role. Without information technology, qualitative analysis of scientometric indicators would be impossible (Shershun, 2020). As a means of combating the great information crisis in the mid-1940s, which was caused by a sharp increase in the amount of scientific information, and when it became impossible to further search for the same information, scientometric databases were created by traditional methods.

The main problem of this subject area is the lack of uniqueness of the publication (Goerlandt & Li, 2021). With the advent of information technology in the field of scientometrics, of course, there has been direct access to the publishing activities of scientists, but on the other hand, this has led to poor science, as well as more frequent plagiarism.

Due to the presence of a huge number of different areas of research in science and an incredibly large amount of information circulating in them, this information is quite difficult to group. That is why it took a lot of time for scientists, in the days before the creation of scientometric databases, to find publications in his field of research and get acquainted with them. The creation of scientometric databases and international registers of stakeholders overcame the information crisis and provided researchers with a convenient mechanism for finding scientific literature and their authors, which allows communication between them and saves time that can be devoted to research (Daradkeh et al., 2022). But it is still a fact that not all researchers use these information technologies developed in the field of scientometrics. This prevents scientists from fully realizing their potential and slows down the development of the scientific field. This problem can be solved by mandatory profile, namely the reflection of scientific activities in scientometric resources, entering the results of their research in international databases. This will give all scientists the opportunity to get acquainted with the work of scientists and evaluate it on scientometric indicators.

Usually, scientific institutions should monitor the process of using scientific databases by scientists, but today not all organizations have a regulated system of activity reporting. Creating a management system for monitoring scientometric components will help in this tracking.

When an issue of evaluating any scientific manuscript arises, it is impossible not to mention Goodhart's law, which has been phrased by Marilyn Strathern as: "When a measure becomes a target, it ceases to be a good measure" (Goodhart's, 2018). The main issue for research is losing the scientific basis when the struggle for finding impact indicators begins. Therefore, it is expedient and relevant to create a multilevel and multi-component assessment of the activities of scientists and institutions. However, creating a system for evaluating the scientific activity of researchers is a complex and painstaking process.

The focus of the topic discussed by some researchers (Van Eck et al., 2014) is present in three popular types of bibliometric networks, namely co-authorship, bibliographic links and joint citation networks, but these ideas extend to other types of bibliometric networks. Two examples of situations with the choice between full and fractional methods are given (Van Eck et al., 2014). Another example concerns bibliographic connected networks of journals. In both examples, it is argued that the fractional counting method is better than the full counting method (Van Eck et al., 2014).

In the context of calculating bibliometric indicators, the concept of publication and co-authorship plays a key role in distinguishing between full and fractional counting. A full count means that the co-authored publication is counted with a total weight of "1" for each co-author, which means that the total weight of the publication is equal to the number of authors of the publication. Fractional counting means that a co-authored publication is assigned fractionally to each of the co-authors, with the total weight of the publication being one (Van Eck et al., 2014).

The idea of visualization of bibliometric networks, which is often called "scientific mapping", attracts serious attention from the first days of bibliometric research (Waltman et al., 2010). Visualization has proven to be a powerful approach to analyzing the wide variety of bibliometric networks, from citation relationship networks between publications or journals to

research collaboration networks or keyword relationship networks. Over time, researchers began to analyze larger and larger networks, which led to the need for more advanced imaging techniques and tools. At the same time, professional bibliometrics users, such as research institutions, financial agencies and publishers, have become increasingly interested in visualizing bibliometric networks. To make visualizations of bibliometric networks available to the general public, both inside and outside the library community, researchers have developed a number of software tools, most of which are freely available (Waltman et al., 2010).

The study (Van Eck et al., 2010) is organized as follows. First, a proposal for a unified approach to mapping and clustering is presented. Then it discusses how the proposed approach relates to previous work published in the physical literature. The following application of the proposed approach is illustrated by creating a combined display and clustering of frequently cited publications in the field of informatics (Van Eck et al., 2010).

There are two types of maps that are commonly used in bibliometric research (Govind et al., 2017). These types of maps are based on distances and graphs. Distance-based maps are maps in which the distance between two elements reflects the strength of the relationship between objects. A shorter distance usually indicates a stronger connection. In many cases, the elements distributed on the maps based on the distance are quite uneven. On the one hand, this makes it easier to identify clusters of related elements, but on the other hand, it is sometimes difficult to mark all the elements on the map without having overlapping labels. Graph-based maps are maps in which the distance between two objects should not reflect the strength of the relationship between the elements. Instead, line points are drawn between them to indicate the relationship. Elements are often fairly evenly distributed on maps based on graphs (Govind et al., 2017).

From the point of view of the analysis of U-I ICT research, this type of meta-analysis must be repeated over time (Olszewska, 2017). Further research would allow a deeper analysis of the elements of the U-I ICT process. It would be particularly interesting to conduct a research to documenti the stages of the activities carried out in this process. It would also allow to compare channels, motivations, activities and outcomes, as well as to more systematically influence U-I ICT factors with the characteristics of CT within national boundaries. Another area of research is the study of institutional differences in international cooperation. The location of the most important factors that must be present can prevent international cooperation, which will end in failure (Olszewska, 2017).

According to Salskyi (Salskyi et al., 2018), the general conclusion is that in the field of publications related to quality management, there are some trends. It should also be noted that the flow of publications related to quality management solutions in industry still exists. However, there are some notable trends in this area as well. Here, the place has evolved from almost dominant production control to leadership, teamwork and customer satisfaction analysis. It is worth mentioning the relevant context of certification, which is gradually declining. In recent years, the link between quality management and a wide range of services is certainly one of the most significant.

3. Experiment and results

Given that the analysis of data is the data of scientists, numerical indicators of their scientific achievements, not clear a preliminary analysis was performed, which noted the need to separate the data in the table for better analysis. *.csv was chosen as the table format. *.csv is a file format that is a delimited format for presenting tabular data, in which fields are separated by commas and newlines. Fields containing commas, multiple lines, or quotation marks (denoted by double quotation marks) should be bounded by quotation marks on both sides. The *.csv format is used to transfer data between databases and spreadsheet editors, which allows you to process data directly from the database.

For the analysis, the table "Scientists" was chosen as the base table, which contains such fields as ID of the scientist, surname in Ukrainian, name in Ukrainian, patronymic in Ukrainian, surname in English, name in English Language, Email, ORCID Identifier, Publons Identifier,

Scopus Identifier, H-Index Scopus, Google Scholar Identifier, Google Scholar H-Index, Number of Publons Publications, Number of Scopus Publications, Number of Google Scholar Publications, Title, Degree, Position, Department , working status, update date, profile ID, draft, h-index Publons. The selected instrumental model shown in Figure 2 indicates not only the ability to work with tables in * .csv format, but also the ability to merge them in the necessary combinations. The "Departments" table was chosen as the second table, containing such fields as ID of the department, name of the department, abbreviation, faculty.

Constructing a probability distribution to represent a data set where identical points are more likely to be selected and dissimilar points have a lower choice. Python is usually used for this algorithm, two Python libraries are needed for construction: pandas, numpy, sklearn and matplotlib.

The following code is used for installation:

```
pip install pandas
pip install numpy
pip install sklearn
pip install matplotlib
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
від sklearn.datasets import load_wine
від sklearn.preprocessing import StandardScaler
від sklearn.manifold import TSNE
```

The sklearn TSNE () class comes with a list of hyperparameters that can be configured when using this method. This is the parameter for measuring the lower space in which you want to convert your data set. In this case, the value must be 2, because it is necessary to visualize the data set in two-dimensional space.

perplexity: float

This is the parameter of the number of nearest neighbors, based on which t-SNE will determine potential neighbors. As a rule, the larger the data set, the greater the value of the surprise:

```
tsne = TSNE (random_state = 0)
tsne_results = tsne.fit_transform(df)
tsne_results=pd.DataFrame(tsne_results, columns=['tsne1', 'tsne2'])
plt.scatter(tsne_results['tsne1'], tsne_results['tsne2'], c=wine.target)
plt.show()
```

The principal components method (PCM) is very productive, but it often fails because it provides the ability to linearly model data. It expresses new features as linear combinations of existing ones, multiplying each by a factor. In order to overcome the limitations of PCM, a number of application techniques have been developed with different characteristic structures. However, diversity-based learning aims to find a method that can cover all data structures.

In different structures, the data have different relationships with each other. For example, they can be linearly separated or very sparse. The relationship of some data to others can be tangent, parallel, spanning or orthogonal. PCM works well with very specific subtypes of data structures because it assumes their initial linearity.

To put this in context, look at 300 x 300 pixel images of the head. Ideally, each image will be perfectly centered, but you really need to consider many additional conditions, such as lighting or facial tilt. If we consider a snapshot of a head as a point in 90,000-dimensional space, changing effects such as tilting the head or the direction of gaze will shift it in space nonlinearly, even though it will remain the same object of the same class.

In real datasets, this type of data is very common. In addition to the described effect, PCM is confused in obtaining distorted representations, extremes, and many imaginary variables in direct unitary coding. In this case, there is a need for a generalized method of reducing the dimension.

Diversity-based learning is relevant to this task. Within this type of training, there are many approaches that you have probably already encountered. These include t-SNE and local-linear attachment (LLA). While PCM seeks to create multiple linear hyperplanes to represent dimensions, much like constructing regressions to estimate data, diversity-based learning seeks to teach diversity represented by smooth, curved surfaces within multidimensional space.

Because such varieties are modeled not by one mathematical function, but by several small linear patches, such linear neighbors are able to model any variety. Although the process of modeling diversity by specific algorithms may be obvious, it is based on the principles themselves.

The following are fundamental aspects of diversity-based learning algorithms:

- There are nonlinear relationships in the data that can be modeled using a variety of smooth and not particularly curved (not very complex) surfaces spanning multiple dimensions. Diversity is continuous.

- It is not necessary to maintain a multidimensional data format. Instead of "compacting" or "projecting" them (as in the case of PCM) according to certain directions to preserve the overall shape, you can perform more complex manipulations, such as deploying a twisted strip or turning the sphere outward.

- The best way to model diversity is to represent a curved space made up of several adjacent elements. If each data point can keep the distance not to other points, but only to the nearest to it, then the geometric relationships in such data can be maintained.

This idea is easier to understand by considering different approaches to deploying a twisted data set.

PCM components: Orange always runs t-SNE on the main input components. This parameter controls the number of key components that should be used when calculating distances between data points.

Data normalization: can apply standardization before starting PCM. Standardization normalizes each column by subtracting the mean of the column and dividing by the standard deviation.

Taking into account these features, two models were obtained:

1. Direct module without additional modules (Figure 1).

2. The model is connected through the method of the nearest neighbor (Figure 2).
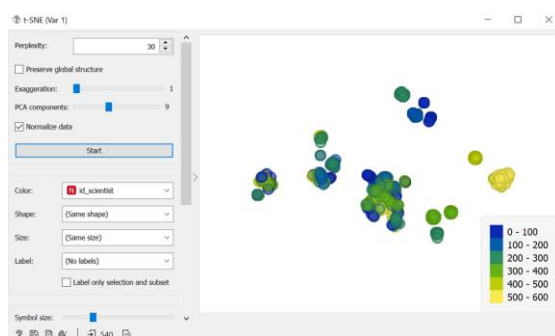

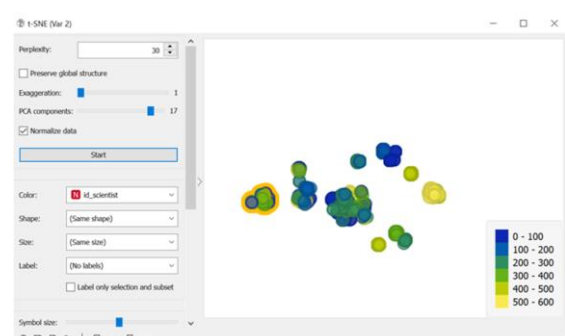**Figure 1.** Direct module without additional modules

Figure 2. The model is connected through the method of the nearest neighbor

If you analyze component data Figure 1 and Figure 2 it can be noted that:

- The number of input and output parameters were equal, which corresponds to the identity of the models.

- The PCM component is distinct. This is primarily due to the direct way it is built. The principal component method (PCM) is a method of factor analysis in statistics that uses

orthogonal conversion of a set of observations with possibly related variables (entities, each of which acquires different numerical values) into a set of variables without linear correlation called principal components.

- The principal component method is one of the main ways to reduce the dimensionality of data by losing the least amount of information. Invented by Carl Pearson in 1901 and supplemented and expanded by Harold Hotelling in 1933. It is used in many fields, including econometrics, bioinformatics, image processing, data compression, and the social sciences.

- The calculation of the principal components can be reduced to the calculation of the singular decomposition of the data matrix or to the calculation of the eigenvectors and eigenvalues of the covariance matrix of the initial data.

Despite or neglecting the numerical indicator of PCM, it can be noted that the overall visualization is almost equivalent.

From a mathematical point of view, the principal components method (Figure 3) is an orthogonal linear transformation that displays data from the original feature space into a new space of smaller dimension.

The first axis of the new coordinate system is built in such a way that the variance of data along it would be maximum. The second axis is built orthogonally to the first so that the variance of data along it would be the maximum possible, etc. The first axis is called the first main component, the second - the second and so on.
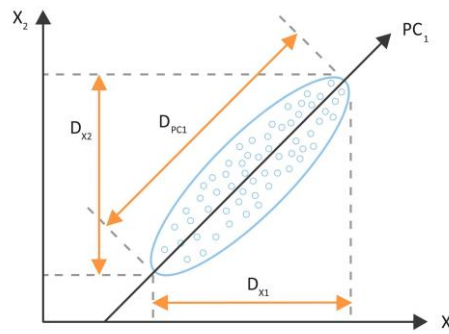

Figure 3. Method of principal components

The figure shows the reduction of the dimension of the original 2-dimensional space using the method of basic components to 1-dimensional. The first main component is directed along the direction of maximum elongation of the ellipsoid of scattering points of objects of the original data set in the feature space, i.e., the greatest variance is connected with it.

In the figure, it is also easy to see that the projection of the data variance on the axis of the first main component is greater than its projection on the original axes but less than their sum. That is. with the help of the first main component it was not possible to express all variance of data. Therefore, build the second, third and so on. The main components until they reflect the total variance. Thus, the meaning of the method is that each main component is associated with a certain proportion of the total variance of the original data set (it is called the load). In turn, the variance, which is a measure of data variability, can reflect the level of their informativeness.

In fact, along some axes of the original space of signs variability may be large, along others - small, and along the third will not be at all. It is assumed that the smaller the variance of the data along the axis, the less significant the contribution of the variable with this axis and, therefore, excluding this axis from space (i.e. variable from the model), you can reduce the dimension of the task almost without losing informative data.

Therefore, the task of the principal components method is to construct a new feature space of smaller dimension, the variance between the axes of which will be redistributed so as to maximize the variance on each of them. To do this, follow these steps:

The total variance of the original feature space is calculated. This cannot be done by simply summing the variances for each variable, as they are usually not independent. Therefore, we need to sum up the mutual variances of the variables determined from the covariance matrix.

The eigenvectors and eigenvalues of the covariance matrix are calculated, which determine the directions of the principal components and the magnitude of the associated variance.

The dimension is reduced. The diagonal elements of the covariance matrix show the variance according to the original coordinate system, and its eigenvalues according to the new one. Then dividing the variance associated with each major component by the sum of the variances for all components, we obtain the proportion of variance associated with each component. After that, so many main components are discarded that the share was 80-90%.

It should be noted that the directive approach to the choice of the number of components does not always give good results. This is because part of the data variance may be due to noise rather than informativeness of the components. Then, setting a threshold of, say, 80%, it may turn out that only 60% of the variance is related to informativeness, and 20 to noise. Therefore, it is almost common to use various special criteria for determining the number of components, such as the Kaiser test, the broken cane test, etc.

The main limitations of the method of basic components are: the impossibility of semantic interpretation of components, because they "absorb" the variance of several source variables. The method can only work with continuous data. The principal components method is included in most analytical platforms and is widely used to reduce the dimensionality of input data at the stage of their processing. The method is sometimes seen as part of a more general approach to reducing the dimensionality of factor analysis data. In analytical platforms in the modules of factor analysis, the method of basic components is often practically implemented.

The database is designed in the PostgreSQL database. It is freely distributed and meets SQL standards as much as possible. PostgreSQL or PostgreSQL try to make full use of ANSI / ISO SQL standards in a timely manner with the release of new versions (Shershun, 2020).

The database consists of 14 tables that store all the necessary information for the full functioning of the system. The base is reduced to the 3rd normal form, contains many connections to many, to each other, and one to many.

The table of scientists stores data on/regarding scientists. It contains the fields of the full name in two languages, e-mail, academic degree, working status, title, department, position, Google Scholar / Scopus / Publons / ORCID profile identifiers related to this scientist's profile, scientometric data (Hirsch indexes, number of publications Google Scholar / Scopus / Publons), update date, profile ID, and draft.

The web application is developed on the Django framework, which is based on Python 3. Django is a free framework for web applications, it uses the MVC architectural template. Django's functionality is enhanced by plug-ins - software modules that allow you to add the necessary functionality to the application. Django also provides security for web applications, namely protection against deliberate attacks in the form of cross-site query forgery (CSRF), XSS attacks and SQL injections (Shershun, 2020).

The DBM application is powered by Django ORM, which connects the database to the concept of object-oriented programming languages, creating a "virtual object database" (Shershun, 2020).

Thanks to the Django ORM library, Python classes are linked to database tables, which does not allow the developer to write SQL queries himself, in a similar manner to the ORM library which builds and executes queries itself. Django ORM reduces application development time, automates the generation of SQL queries, provides the ability to change the database in the representation of object-oriented programming languages, performing SQL updates when changing / adding / deleting object fields.

Navigation is an important part of a web application. In the Django framework, navigation is defined in the urls.py file. The URL configuration is pure Python code and is displayed between URL path expressions and Python functions (Salskyi et al., 2018).

First of all, we connect libraries to work with the administrative panel, accounts and functions of displaying HTML templates.

```
from django.contrib import admin
from django.contrib.auth.decorators import login_required
from django.urls import path, include
from app import views
urlpatterns = (
    path('admin/', admin.site.urls),
    path('accounts/', include('django.contrib.auth.urls')),
    path('', views.MainPage.as_view()),
    path('scientists/', views.ScientistsPage.as_view(), name="scientists"),
    path('scientists/search', views.Search.as_view(), name="search"),
    path('profile/<str:profile_id>', views.ProfilePage.as_view(), name="profile"),
    path('information/', views.information, name="information"),
    path('report/', login_required(views.report), name="report"),
    path('report/export_xls/', login_required(views.export_xls), name="export_xls"),
```

The variable "urlpatterns" contains the sequence of paths django.urls.path () and / or django.urls.re_path () instances. The first path is responsible for accessing the administrative panel, the next - for the path for authorization in the system. The third line calls the "MainPage" function from the views file, which displays the main page template. The "scientists" path displays a page with a list of all scientists, as well as the ability to search for "scientists / search". The next path responsible for displaying the scientist's profile parameter "<str: profile_id>" determines that the unique identifier for each scientist has a data type - String. Thanks to the "login_required ()" method by "report" or "report / export_xls" - the user can follow the link only when he is authorized in the system, this method checks the flag "is_authenticated".

An important function of a web application is to generate reports based on data in the system. The following function describes how reports are generated.

The first line is a decorator and indicates that this function occurs only when the user is authorized. Then begins the function "export_xls", which specifies the file, namely ".xls", as well as its name - "Report.xls". Next, the entire file is assigned the UTF-8 encoding, as well as the name for the first letter. The cells of the first line are set styles and their name is specified.

As a next step, the styles for all cells are announced, according to scientists. The variable "rows" records an array of scientists, namely their data, which are necessary for reporting and are sorted alphabetically by last name. Thanks to the cycle, the table is filled with data about scientists. The last step is to set the width of all columns and save the file, after which the file is downloaded by the client.

Recommendations for implementation and use. The software developed as part of the research has a classification of a client-server application with a three-tier architecture: client, application server and database server. This requires the following optimal system requirements for Google Chrome starting with version 83.0.4103.61. Any browser program is required to use this application, but it is recommended to use Google Chrome, because on this browser the application was tested. The system requirements for optimal use of the browser are presented below.

4. Conclusion

It was found (Salskyi et al., 2018), that the main problem of the subject area is the systematization and further processing of the obtained amount of data, which over time approaches the standards of Big Data. Usually scientometric data include identifiers of scientists, indicators

such as h-index, SJR, SNIP, IF. All these data should not only be stored for each institution for each specific NNP, but also to form reporting and rating documentation.

Given the fact that the above analysis of the data are the data of scientists who were separated into appropriate tables for better analysis, * .csv was chosen as the table format.

For the analysis, the table "Scientists" was chosen as the base table, which contains such fields as ID of the scientist, surname in Ukrainian, name in Ukrainian, patronymic in Ukrainian, surname in English, name in English Language, Email, ORCID Identifier, Publons Identifier, Scopus Identifier, H-Index Scopus, Google Scholar Identifier, Google Scholar H-Index, Number of Publons Publications, Number of Scopus Publications, Number of Google Scholar Publications, Title, Degree, Position, Department , working status, update date, profile ID, draft, h-index Publons. The selected instrumental model is shown in Figure 2 shows not only the ability to work with tables in *.csv format, but also to combine them in the necessary combinations. The table "Departments" was chosen as the second table, which contains such fields as ID of the department, name of the department, abbreviation, faculty.

Considering the existing needs of the system, the requirements for the developed product were formulated. Based on modern technologies and approaches to program development, the general steps for the implementation of the system were formulated. The result is a full-fledged web application that allows convenient and remote access to scientometric information of scientific and pedagogical staff of ONUT.

Developed as an appendix, as part of a research project, designed to improve the formation of reporting documentation and ranking of ONUTs' scientists with the involvement of standards of stakeholders in the field of scientometrics. To achieve this goal, the main opportunities and benefits of the standards set by the world scientific community were carefully studied and analyzed.

Assessment methods (ranking) are divided into traditional and non-traditional. Traditional methods are effective in large hierarchical organizations that operate in a fairly stable environment. Their disadvantages are that the assessment is given to an individual employee without taking into account the goals of the organization, based on the assessment of the head, while completely ignoring the opinion of colleagues, subordinates, clients, etc.; focuses on the past (achieving results) and does not take into account the long-term prospects of the organization and the employee. Non-traditional methods consider the team and emphasize the assessment of the employee by his colleagues and the ability to work in a group.

Creating a high-level assessment of a researcher's scientific activity is a complex and painstaking process. Every week, or even daily, each researcher has to fill in and generate an infinite number of reports, which should include: number of publications (total and in Scopus and / or Web of Science Core Collection, respectively), h-index, number of citations, etc. However, each of these factors composes many characteristics.

First, the application of such a parameter as the number of citations to the assessment of the scientist, cannot perfectly characterize the quality of scientific work. The worst publication can be cited many times, which will raise the rating of the scientist, but will not reflect the quality of scientific work, but will increase the h-index.

Secondly, the value of h-index cannot be used as an average value of a scientist's activity because the citation rate of a researcher in a narrow profile cannot be compared with a scientist working in the field of knowledge, which is common among a large number of researchers.

To achieve the best quality result, a careful study of the presented analogues was conducted, based on this study it was found that not one of the presented analogues does not fully meet the basic requirements and needs of the target audience. This resource realizes the main advantages of world standards of recognized stakeholders in the industry.

A software product that meets all the basic needs of users, thanks to a convenient and clear design, proper structuring of information presented, organized web application navigation using animation and the latest design standards, software support and constant updating of final display data for higher education and staff of the academy.

In the context of research, "scientific information", according to the European Commission, means peer-reviewed scientific articles published in scientific periodicals, research data and other forms of scientific activity, such as monographs and conference proceedings. This "scientific information" in accordance with the basic concept of Science 2.0 is the interaction of the research community and components of open science, including Open Access, Open Code, Citizen Science. When applied to research, this has a huge impact on significantly improving the quality of scientific content. High availability of data increases the ability to engage in science and provides interaction between scientists, which in turn increases the focus on specific scientific problems.

The application of the principles of Science 2.0 causes a huge increase in the number of researchers who are constantly searching for information for their fields of knowledge. Such platforms include the digital online repository figshare, the scientific platforms academia.edu and research gate, the Mendeley abstract manager, the Altmetric and Impact Story portals. All these systems have been identified by the European Commission for Research.

REFERENCES

1. Daradkeh, M., Abualigah, L., Atalla, S. & Mansoor, W. (2022). Scientometric Analysis and Classification of Research Using Convolutional Neural Networks: A Case Study in Data Science and Analytics. *Electronics,* 11(13), 2066.

2. Goerlandt, F. & Li, J. (2021). *Forty Years of Risk Analysis: A Scientometric Overview*. Risk Analysis.

3. Goodhart's law. (2018, July 09). Retrieved July 15, 2018, from https://en.wikipedia.org/wiki/Goodhart's_law.

4. Govind, M., Küttim, M. (2017). International knowledge transfer from university to industry: A systematic literature review. *Research in Economics and Business: Central and Eastern Europe,* 8(2), 5-25.

5. Olszewska, A. (2017). Research Issues Undertaken within Quality Management – Overview of Selected Literature and a Knowledge Map. *Procedia Engineering*, 182, 518–523. doi:10.1016/j.proeng.2017.03.146.

6. Salskyi, D., Kozhukhar, A., Olshevska, O. & Povarova, N. (2018). Overview of popular approaches in creating client-server applications based on scientometrics onafts' platform. *Automation of technological and business processes,* 9(4). doi:10.15673/atbp.v10i4.833.

7. Shershun, O. (2020). Informatsiino upravliaiucha systema obrobky naukometrychnykh danykh profesorsko-vykladatskoho skladu ONAFT. *Realizatsiia kliientskoi chastyny: dyploma*. 14-60. (Ukrainian source).

8. Van Eck, N., Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics,* 84(2), 523–538. doi:10.1007/s11192-009-0146-3.

9. Van Eck, N., Waltman, L. (2014). *Visualizing bibliometric networks*. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), Measuring scholarly impact: Methods and practice, 285-320. Springer.

10. Waltman, L., Van Eck, N., Noyons, E. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics,* 4(4), 629–635. doi:10.1016/j.joi.2010.07.002.

Oleksandr SHERSHUN, IT specialist at the Scientific and Technical Library, Web developer and DevOps-developer at Odessa National University of Technology. He has 4 years of work experience. The total number of publications is more than 12.

Iryna ZINCHENKO, Former-Director of the Scientific and Technical Library of the Odessa National University of Technology, Librarian-Bibliographer with the highest qualifications. Her total work experience in the industry exceeds 30 years. The main scientific direction is project management. The total number of publications is over 80.

Zhanna TYTURENKO, IT specialist of the Scientific and Technical Library, Account and Social Media Mananager of the Scientific and Technical Library of the Odessa National University of Technology. She has 2 years of work experience. The total number of publications is more than 12.

Anastasiia VOLKOVA, IT specialist of the Scientific and Technical Library, Head of Coordinating Centre of Scientific Journals' Publishing of the Odessa National University of Technology, Specialist in the field of data analysis. She has 3 years of work experience. The total number of publications is more than 12.

Olga OLSHEVSKA, PhD, Associate Professor at the Department of Information Technology and Cybersecurity, Director of the Scientific and Technical Library of the Odessa National University of Technology. Her total work experience in the industry exceeds 8 years. Her main research areas are data analysis, project management, software development in accordance with standards open access and paperless. The total number of publications is over 50.