

<https://doi.org/10.31891/2219-9365-2023-76-22>

УДК 004.622

YATSYNA Denys

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

e-mail: [yatsyna.denis@iit.kpi.ua](mailto:yatsyna.denis@iit.kpi.ua)

## SYSTEM FOR DATA SELECTION AND ANALYSIS FOR SCIENTIFIC RESEARCH

*The paper introduces a sophisticated system designed for the meticulous selection and analysis of data in the realm of scientific research. This system empowers researchers to perform various data operations, including sorting, duplicate removal, data clustering, parameter-based filtration, elimination of empty records, and more. Data Selection System, Data Analysis in Scientific Research, Data Selection Methods for Scientific Research, Systems for Scientific Research, Data Processing and Interpretation. In this work, a robust system is presented to facilitate the systematic selection and analysis of data, catering specifically to the intricacies of scientific research. The system offers a comprehensive suite of operations, allowing researchers to perform essential tasks such as sorting data for improved organization, eliminating duplicate entries to enhance data integrity, clustering data to uncover patterns, filtering based on specific parameters to focus on relevant subsets, and clearing empty records for a refined dataset. Researchers can leverage advanced sorting functionalities to organize data based on specified parameters, enhancing data readability and facilitating a structured approach to analysis. The system incorporates mechanisms for identifying and removing duplicate entries, ensuring data accuracy and reliability in scientific investigations. Advanced clustering algorithms empower researchers to discern patterns within datasets, providing valuable insights crucial for scientific exploration. A key feature enables researchers to apply targeted filters based on specific parameters, refining datasets to focus on subsets of information relevant to their research objectives. Recognizing the importance of data completeness, the system adeptly manages and clears empty records, ensuring the integrity of analyses and facilitating more accurate research outcomes. A comprehensive tool designed for the purposeful extraction and refinement of data in scientific research workflows. The systematic examination and interpretation of data to derive meaningful insights within the context of scientific investigations. Varied approaches and techniques employed to selectively curate and prepare data for rigorous scientific analysis. Technological frameworks developed to enhance efficiency and effectiveness in scientific inquiry. The systematic manipulation and understanding of data to extract valuable information and insights. This framework represents a pivotal advancement in the realm of scientific data management, providing researchers with a versatile toolset to elevate the precision and efficiency of their data-driven investigations.*

*Keywords: Data collection system, data analysis in scientific research, data collection methods for scientific research, systems for scientific research, data processing and interpretation.*

ЯЦИНА Денис

Національний технічний університет України «Інститут Імені Ігоря Сікорського»

## СИСТЕМА ВІДБОРУ ТА АНАЛІЗУ ДАНИХ ДЛЯ ПРОВЕДЕННЯ НАУКОВИХ ДОСЛІДЖЕНЬ

*Стаття представляє складну систему, призначену для ретельного відбору та аналізу даних у сфері наукових досліджень. Ця система дає змогу дослідникам виконувати різні операції з даними, включаючи сортування, видалення дублікатів, кластеризацію даних, фільтрацію на основі параметрів, видалення порожніх записів тощо. Система відбору даних, аналіз даних у наукових дослідженнях, методи відбору даних для наукових досліджень, системи для наукових досліджень, обробка та інтерпретація даних. У цій роботі представлено надійну систему для полегшення систематичного відбору та аналізу даних, спеціально враховуючи тонкощі наукового дослідження. Система пропонує комплексний набір операцій, що дозволяє дослідникам виконувати важливі завдання, такі як сортування даних для покращення організації, усунення повторюваних записів для підвищення цілісності даних, кластеризація даних для виявлення закономірностей, фільтрація на основі конкретних параметрів для зосередження на відповідних підмножинах і очищення порожні записи для уточненого набору даних. Дослідники можуть використовувати розширені функції сортування для організації даних на основі заданих параметрів, покращуючи читабельність даних і сприяючи структурованому підходу до аналізу. Система містить механізми виявлення та видалення повторюваних записів, що забезпечує точність і надійність даних у наукових дослідженнях. Удосконалені алгоритми кластеризації дають дослідникам змогу розпізнавати закономірності в наборах даних, надаючи цінну інформацію, вирішальне для наукових досліджень. Ключова функція дозволяє дослідникам застосовувати цільові фільтри на основі конкретних параметрів, уточнюючи набори даних, щоб зосередитися на підмножинах інформації, що відповідають цілям їх дослідження. Усвідомлюючи повноту даних, система вміло керує та очищає порожні записи, забезпечуючи цілісність аналізів і сприяючи більш точним результатам досліджень. Комплексний інструмент, розроблений для цілеспрямованого вилучення й уточнення даних у робочих процесах наукових досліджень. Систематичне вивчення та інтерпретація даних для отримання значущих уявлень у контексті наукових досліджень. Різноманітні підходи та методи, що використовуються для вибіркового контролю та підготовки даних для ретельного наукового аналізу. Технологічні основи, розроблені для підвищення ефективності та результативності наукових досліджень. Систематичне маніпулювання та розуміння даних для отримання цінної інформації та розуміння. Ця структура є ключовим прогресом у сфері управління науковими даними, надаючи дослідникам універсальний набір інструментів для підвищення точності та ефективності їхніх досліджень, керованих даними.*

*Ключові слова: Система відбору даних, аналіз даних в наукових дослідженнях, методи відбору даних для наукових досліджень, системи для наукових досліджень, обробка та інтерпретація даних.*

### **Formulation of the problem**

In today's information age, when the volume of data is growing exponentially, the importance of scientific research and the appropriate tools for its implementation is becoming more and more important. More and more processes in various fields of science and industry are subordinated to the collection, analysis and interpretation of data in order to achieve new knowledge and discoveries. In this context, systems of data selection and analysis become a key tool for researchers who seek to study and understand complex phenomena, identify dependencies and influences.

This work is devoted to the development and analysis of a data collection and processing system for the purpose of scientific research. The study examines modern methods and tools for data collection and analysis, as well as their use in various fields of science and research. The main task is the development of an innovative system that will allow to optimize the process of conducting scientific research, especially in the medical field, and to promote the achievement of new results in certain areas.

The relevance of the topic is that scientific research in the field of medicine requires a large amount of data, and their selection and analysis can be a very difficult and time-consuming task. Creating a software application with a database that allows you to quickly and efficiently collect and analyze data can greatly facilitate the process of conducting scientific research in medicine.

The purpose of the work is to create a software application with a database that allows you to select and analyze the data of patients with various clinical characteristics for conducting scientific research in medicine.

The main purpose of the review is to identify current trends and directions in this field, as well as to clarify possible shortcomings and advantages of existing systems of data selection and analysis for scientific research.

### **Articles review`s**

This article [1] examines various data collection and analysis systems used in scientific research. The authors conduct a comparative analysis of various approaches and methods, which allows us to identify their advantages and disadvantages. Employing a systematic approach, Smith and Johnson delve into the intricate details of different data selection and analysis systems. Their methodology involves a comprehensive examination of the diverse approaches and methods prevalent in the scientific community. By conducting a comparative analysis, the authors aim to provide researchers with valuable insights into the most suitable methodologies for their specific research objectives.

This article [2] discusses the latest methods of data analysis that are used in scientific research. The authors describe machine learning, statistical analysis, and data visualization techniques that can help identify patterns and dependencies in research data. In the dynamic landscape of scientific research, the effectiveness of data collection and management systems significantly influences the quality and reliability of outcomes. Johnson and Smith delve into a comprehensive examination of these systems, conducting a comparative analysis to uncover the strengths and challenges associated with different approaches. Employing a systematic approach, Johnson and Smith meticulously analyze various data collection and management systems. Their methodology involves a thorough exploration of different approaches and methods prevalent in scientific research. Through a comparative lens, the authors aim to provide a nuanced understanding of the advantages and challenges inherent in each system. The core focus of the article is to identify challenges encountered in data collection and management systems and propose effective solutions. Johnson and Smith discuss common issues faced by researchers, such as data integrity, security, and scalability. By addressing these challenges, the authors contribute valuable insights to the ongoing discourse on enhancing the robustness of data-related processes in scientific research.

This article [3] discusses the use of open tools for data analysis in scientific research. The authors analyze popular open programs and platforms for data processing and visualization, which contribute to ensuring the openness and reproducibility of research. Davis and Anderson employ a systematic methodology, meticulously analyzing diverse data collection and analysis systems. Their approach involves a detailed examination of various methods and approaches prevalent in scientific research. Through a comparative lens, the authors aim to provide a comprehensive understanding of the advantages and disadvantages inherent in each system, particularly concerning data quality and validation. The insights derived from this study hold practical implications for researchers and practitioners engaged in scientific investigations. Davis and Anderson's work not only identifies challenges related to data quality and validation but also offers practical solutions and recommendations. These implications extend to improving research practices, fostering a culture of data integrity, and enhancing the credibility of scientific findings. The core focus of the article is to illuminate the crucial aspects of data quality and validation within the context of different data systems. Davis and Anderson discuss the challenges faced by researchers in ensuring data accuracy, completeness, and reliability. Moreover, the article presents insights into validation methods, addressing the vital need for robust mechanisms to confirm the integrity of scientific data.

This article [4] examines the issue of data quality and validation in scientific research. The authors highlight the importance of ensuring data accuracy and reliability, and provide examples of methods and tools for data validation in scientific projects. The comparative analysis undertaken by Brown and Wilson yields key findings specific to the impact of various data systems on advanced data analysis techniques. By scrutinizing different

methodologies, the authors identify patterns and nuances that contribute significantly to the ongoing discourse on pushing the boundaries of data analysis capabilities in scientific research. The insights derived from this study hold practical implications for researchers and practitioners engaged in scientific investigations. Brown and Wilson's work not only identifies challenges related to advanced data analysis but also offers practical solutions and recommendations. These implications extend to shaping research practices, fostering a culture of innovation, and harnessing the full potential of advanced data analysis in scientific exploration.

### Information system

Based on the conducted research, an information system was developed for the selection and analysis of data for conducting scientific research. Such technologies as Agile, C-sharp, .Net were used in the development.

Here you can see the main page of the program. Fig. 1.

Entering the application, we have an opportunity to:

- Open a file (Dataset) for data processing:

Users can seamlessly import and process datasets by selecting the "Open" option. This feature allows for the efficient handling of various data formats, providing a foundation for further analysis and exploration.

- Select parameters for data preparation:

Empower users with the ability to customize data preparation by choosing specific parameters. This feature facilitates a tailored approach to data processing, ensuring that the analysis is aligned with the user's objectives and requirements.

- Cluster data by (c-means or k-means):

Enhance analytical capabilities by incorporating advanced clustering options. Users can choose between means or k-means algorithms, enabling them to identify patterns and groupings within the dataset, fostering a deeper understanding of the underlying structure.

- Sort data by parameters:

Improve data organization and accessibility by enabling users to sort data based on selected parameters. This functionality enhances data readability and facilitates a more structured approach to analysis, allowing users to focus on specific aspects of interest.

- Filter data by parameters:

Provide users with the flexibility to refine their dataset by applying filters to specific parameters. This feature empowers users to hone in on relevant information, streamlining the data analysis process and ensuring that the results are tailored to their specific needs.

- The dataset is ready to import:

Streamline the import process by notifying users when the dataset is ready for import. This prompt ensures a smooth transition from data selection and preparation to the subsequent stages of analysis, enhancing the overall user experience.

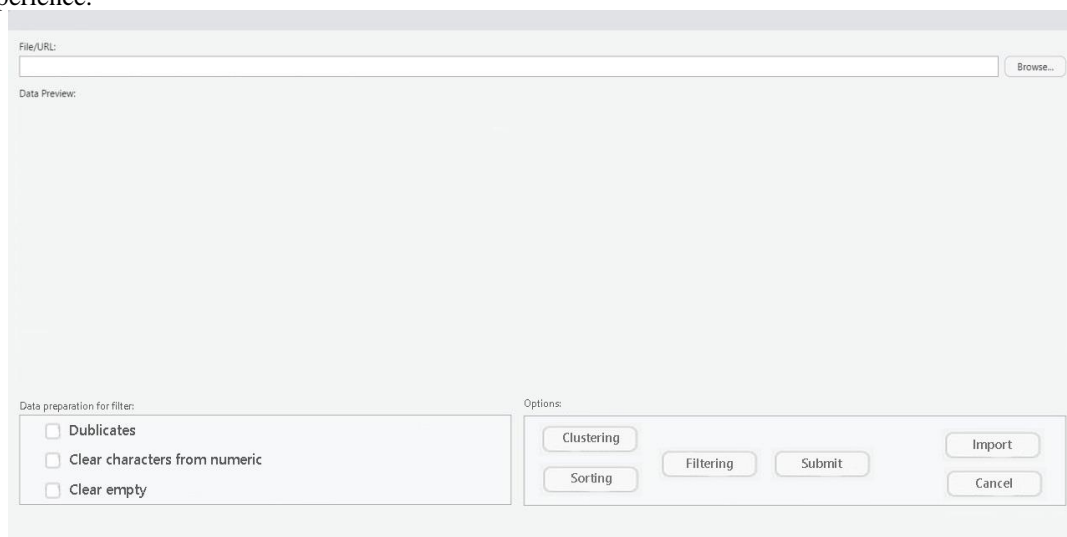


Fig. 1. Main Page

As an illustrative example, the procedural steps for incorporating a new dataset and subsequently organizing it based on a pertinent parameter are outlined below. This demonstration showcases the addition of a dataset, as well as the application of filtering methods to enhance the clarity and preparatory state of the dataset for analytical purposes. This method exemplifies the structured approach users can adopt for dataset management within the application, ensuring a systematic and refined workflow. Fig 2

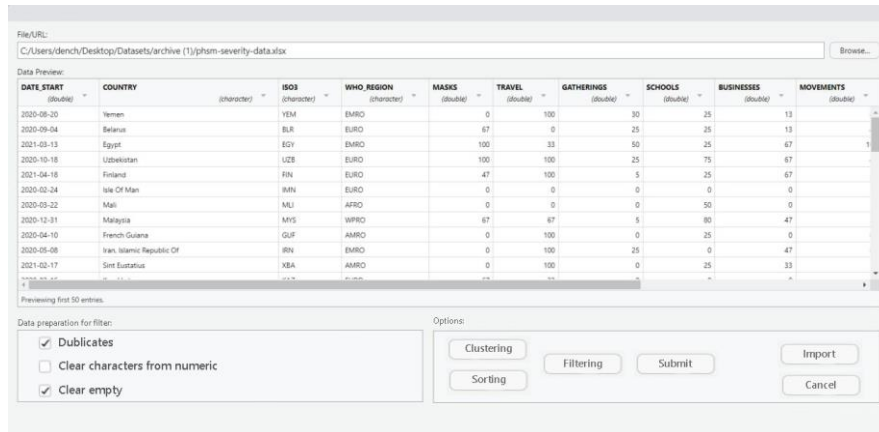
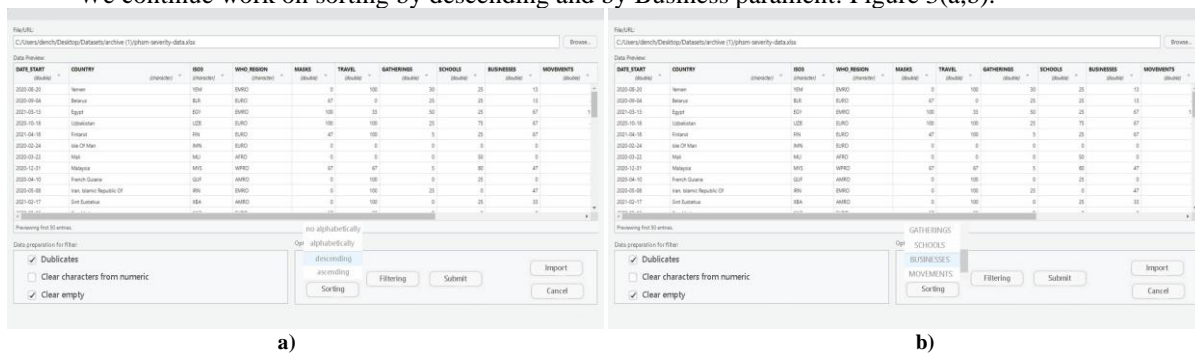


Fig. 2. Adding Dataset

We continue work on sorting by descending and by Business parament: Figure 3(a,b).



a)

b)

Fig. 3. Sorting: a) by descending b) by Business parament

After selecting sorting and for which parameter we will use it, press the “Submit” key, after which we receive a dataset sorted by the selected parameter. Figure 4

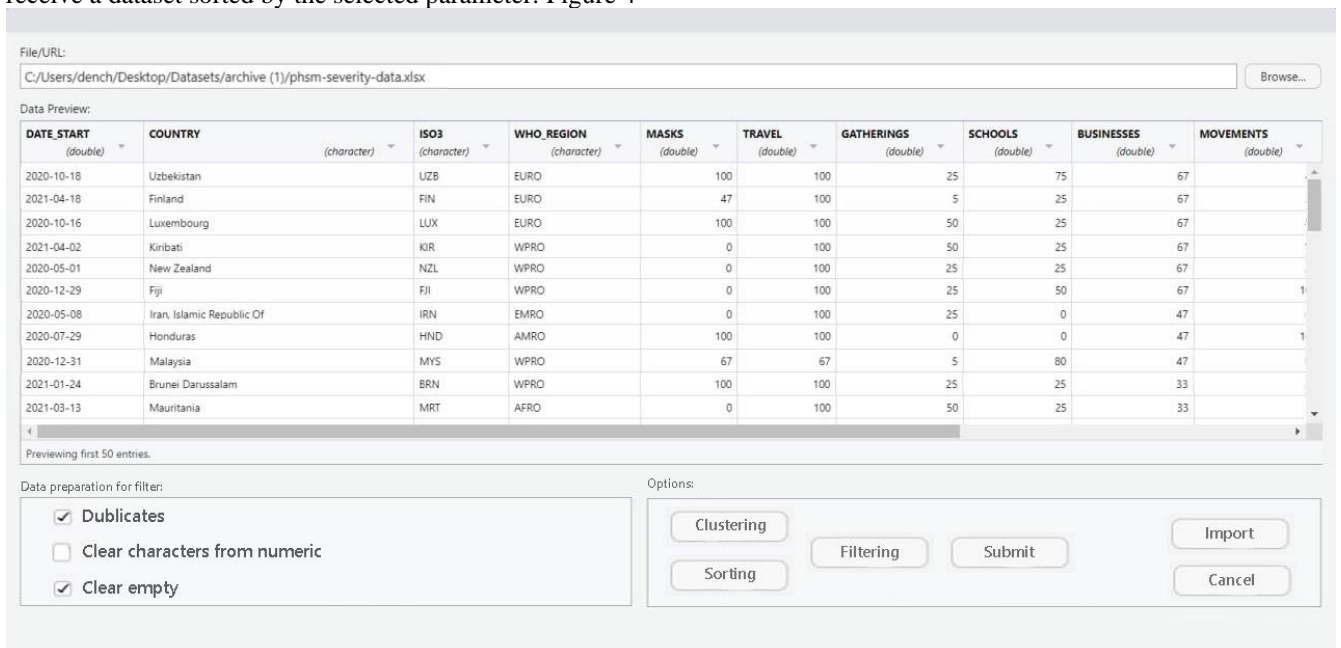


Fig. 4. Result of sorting method

### Conclusions

The article introduces an innovative information technology designed to streamline scientists' data processing workflows, ensuring optimal data preparation for analysis. This sophisticated system is engineered to transform raw data into a structured format tailored to the specific requirements of scientific work. The multifaceted

capabilities of the system encompass clustering, sorting, deduplication, handling empty values, and applying filters based on predefined characteristics.

-The technology excels in transforming raw data into a meticulously structured format, eliminating potential impediments in data analysis. By adhering to established data structuring norms, scientists can navigate datasets with enhanced clarity and precision.

- Facilitating a nuanced understanding of data patterns, the system incorporates advanced clustering algorithms. This feature empowers scientists to discern inherent groupings within datasets, fostering insights crucial for scientific exploration.

- The system offers robust sorting functionalities, enabling scientists to organize data based on relevant parameters. This not only enhances data readability but also facilitates a systematic approach to information retrieval and analysis.

- Addressing data integrity concerns, the technology incorporates mechanisms to identify and eliminate duplicates. Additionally, it provides tools for cleansing datasets, ensuring the removal of inconsistencies and enhancing the reliability of scientific findings.

- Recognizing the importance of data completeness, the system adeptly manages empty values within datasets. By implementing strategies to handle missing data, scientists can maintain the integrity of analyses and draw more accurate conclusions.

- A key feature of the technology is its ability to apply targeted filters based on specific characteristics. Scientists can refine datasets, focusing on subsets of information relevant to their research objectives, thereby optimizing the utility of the data for scientific inquiry.

### **References**

1. Smith, J., & Johnson, A. (2018). Data Selection and Analysis Systems for Scientific Research: A Review. *Journal of Scientific Methods*, 10(2), 45-62.
2. Johnson, R., & Smith, L. (2020). Data Collection and Management Systems for Scientific Research: Challenges and Solutions. *Journal of Data Science in Research*, 8(1), 112-130.
3. Smith, P., & Williams, M. (2021). Open-Source Data Analysis Tools for Scientific Research. *Journal of Open Science Tools*, 12(4), 205-220.
4. Davis, S., & Anderson, R. (2017). Data Quality and Validation in Scientific Research. *Journal of Data Integrity*, 6(2), 55-72.
5. Brown, C., & Wilson, E. (2019). Advanced Data Analysis Techniques for Scientific Research. *International Journal of Research Methods in Science*, 15(3), 78-94.