

Алла Федорівна Дасів,*канд. екон. наук*

Інститут економіки промисловості НАН України

вул. Марії Капніст, 2, м. Київ, 03057, Україна

E-mail: alladasiv@gmail.com<https://orcid.org/0000-0001-5431-701X>;**Олена Анатоліївна Руссіян,***канд. екон. наук, старший науковий співробітник*

Інститут економіки промисловості НАН України

вул. Марії Капніст, 2, м. Київ, 03057, Україна

E-mail: elena.russiyan@gmail.com<https://orcid.org/0000-0002-1246-9952>;**Денис Володимирович Липницький,***канд. екон. наук, провідний інженер із розроблення ПЗ*

Vodafone Germany GmbH, м. Дюссельдорф, 40001, Німеччина

ФОРМУВАННЯ БАЗИ ТА АВТОМАТИЗОВАНОЇ СИСТЕМИ ІНТЕГРАЦІЇ ДАНИХ ПРО ПРОМИСЛОВІСТЬ¹

У статті обґрунтовано доцільність формування спеціалізованої бази даних про промисловість і системи автоматизованого їх збору та використання. Ця система дозволяє забезпечити інформаційні потреби аналітиків і науковців щодо проблем розвитку промисловості. Вона передбачає автоматизований збір інформації за багатьма країнами світу з використанням будь-якої кількості джерел, доступних як клієнт-серверні ресурси в інтернеті. Застосування сучасних алгоритмів інтеграції даних забезпечує різноманітну подачу інформації, формати її надання та періодичність оновлення. Крім того, просунутим користувачам системи надається широке коло можливостей для конструювання критерію пошуку та формату отримання даних. Система побудована як клієнт-серверна технологія та є доступною для безшовної інтеграції з аналогічними системами як джерело та сховище інформації.

Для побудови системи виконано аналіз існуючих у світі та популярних у використанні статистичних баз даних, визначено їх особливості, переваги і недоліки. Встановлено, що в більшості існуючих статистичних систем не представлено статистичних даних по Україні, а там, де вони є, відсутня деталізація даних за видами економічної діяльності та галузями, особливо по промисловості. До того ж мають місце певні технічні труднощі в роботі з даними для користувачів, персоналізація доступу майже відсутня. Досить обмеженими є можливості інтеграції існуючих баз даних із системами, що використовують інформацію для моделювання та прогнозування як у режимі запиту, так і в режимі реального часу.

Сформовано структуру статистичної бази даних про розвиток промисловості, відібрано показники для її наповнення. Показники розподілено за секторами економіки (сектор виробництва сировини, переробний сектор, макростатистика, інші сектори). Склад секторів розкрито за ключовими видами економічної діяльності відповідно до класифікації видів економічної діяльності 2010 року (КВЕД-2010), які забезпечують переважну частину створення доданої вартості економіки України. Кожен визначений вид економічної діяльності розкрито за групами галузей та галузями. Крім того, базу статистичних даних сформовано на основі призначення кожному з показників інших характеристик, необхідних для автоматизації: код, одиниці вимірювання, період, база (розподілення показників за окремими гру-

¹ Стаття підготовлена в рамках виконання конкурсної тематики «Смарт-система промислового прогнозування і форсайтингу» (номер держреєстрації 0122U002165).

пами для можливості моделювання), джерело (статистичні бази та джерела, з яких здійснювався збір показників і даних). Для зберігання інформації використано сучасну реляційну базу даних, яка дозволяє вирішувати питання оптимізації для роботи з найбільш потужними, але ще не великими даними, ураховує особливості широких даних, уможлиблює горизонтальне та вертикальне масштабування, у тому числі в системі баз даних із відкритим вихідним кодом PostgreSQL.

Сформовано методику формування і технологію наповнення статистичної бази даних автоматично з великої кількості джерел, режим доступу до яких задається параметрами конфігурації системи. Розроблено технологію міграції даних, доступних у гнучких форматах, у тому числі текстовому, зокрема Excel.

Використання розробленої автоматизованої інформаційної системи інтеграції промислової статистики дозволяє вибирати необхідні показники для аналізу економічних процесів у промисловості, використовувати зібрані в єдиному інформаційному просторі статистичні дані за 130 країнами світу, у тому числі України, для здійснення наукових досліджень, побудови економіко-математичних моделей і прийняття управлінських рішень.

Ключові слова: промисловість, економічні показники, статистика, база даних, автоматизована система інтеграції даних, інформаційна технологія, економічна аналітика.

JEL: L86, L6, L52, L16, F47, E27, C82, C81

Сучасний етап розвитку інформатизації характеризується як надзвичайно швидким зростанням обсягів інформації, насамперед цифрової, так і наявністю різноманітних джерел її отримання. У сучасних промислових смарт-системах інформація посідає центральне місце (Вишневський, Вієцька, Гаркушенко та ін., 2018). Із використанням великих обсягів даних і штучного інтелекту формуються новітні системи управління промисловістю (Yin, Kaynak, 2015; Zarte, Pechmann, Nunes, 2019; Вишневський, Вієцька, Вієцький та ін., 2019; Li, Chen, Shang, 2022). Як відзначено в роботі (Li, Chen, Shang, 2022), інформаційні технології роблять промисловість більш інтелектуальною. При цьому аналіз, заснований на різноманітних даних, підвищує її ринкову конкурентоспроможність за рахунок вилучення прихованих знань і виявлення потенційних можливостей розвитку виробництва, а також допомагає керівникам підприємств приймати розумні рішення в різних складних виробничих умовах. Але потоки даних набувають особливого значення не тільки на рівні суб'єктів господарювання, але і на рівні окремих галузей (Чесалов, Шпильман, Аракчеев, 2016) і промисловості загалом, де використовуються переважно попередньо оброблені дані у вигляді статистичної

інформації, яку зазвичай будують згідно з принципами реляційних баз (Delplanque, Etien, Anquetil, Auverlot, 2018).

Формування таких баз даних є непростим завданням – воно пов'язане з комплексом різнопланових проблем, таких як складність аналізу та візуалізації залежностей між об'єктами бази даних, оцінки впливу модифікацій на бази даних, тестування їх функціональних можливостей, необхідність забезпечення інформаційної безпеки тощо (Delplanque, Etien, Anquetil, Auverlot, 2018; Полтавцева, Хабаров, 2016).

При здійсненні аналітичних досліджень виникають ускладнення в процесах пошуку, обробки, систематизації даних, зручних для використання та побудови інформативних звітів. Існуючі вітчизняні та зарубіжні статистичні бази надають інструментарій і можливість доступу до інформації, що характеризує динаміку соціально-економічних процесів в економіці. До того ж вони містять деякі дані щодо розвитку промисловості, але їх обсяг і перелік є обмеженими з позиції системних досліджень проблем індустріального розвитку. В Україні на сьогоднішній день не сформовано спеціалізованої статистичної бази даних про промисловість. При цьому тенденції цифровізації промисловості вказують на актуальність завдання розроблення

автоматизованої бази даних розвитку промисловості у вигляді окремої статистичної системи. Вона дозволить зручно користуватися даними та може розглядатися як складовий модуль смарт-системи промислового прогнозування і форсайтингу, яка розробляється в Інституті економіки промисловості НАН України. Зокрема, база таких даних може бути центральним елементом інформаційного забезпечення для комплексу економіко-математичних моделей розвитку промисловості (Дасив, Мадых, Охтень, Турлакова, 2019).

Метою статті є обґрунтування концептуальних положень щодо формування бази і системи автоматизованої інтеграції даних промислової статистики, призначеної для інформаційного забезпечення аналітичних досліджень і прийняття управлінських рішень із використанням сучасного наукового інструментарію.

Структуру статті побудовано таким чином: аналіз характеристики наявних баз статистичних даних; визначення їх переваг і недоліків; порядок формування і технології автоматизованого наповнення спеціалізованої бази даних про національну і світову промисловість. Завершують статтю короткі підсумки та пропозиції щодо подальших напрямів досліджень.

З урахуванням необхідності формування спеціалізованої бази даних про промисловість доцільно проаналізувати відомі у світі статистичні системи, до яких належать: OECD.Stat (Organization for economic co-operation and development), Eurostat Data Browser (European Commission), The World bank DataBank (The World Bank Group), а також систему статистичної інформації Державної служби статистики України.

У результаті аналізу статистичної бази даних OECD Data¹ виокремлено такі її переваги у використанні: зручність дерева вибору показників; інформативність підрозділів; можливість вибору в таблиці необхідного показника, країни, одиниці вимірювання. Обмеження бази полягає в тако-

¹ OECD.Stat (2022). URL: <https://stats.oecd.org/>

му: обирати для дослідження можна статистичні дані тільки за одним показником чи країною, що ускладнює формування порівняльних таблиць. До цієї системи не включено статистичні дані України. Певні незручності також виникають через те, що показники не структуровані за узагальненими групами, а розміщені загальним дуже великим переліком.

Позитивним прикладом побудови структури статистичної бази даних (дерева статистики) є Eurostat Data Browser², що має досить наочну структуру та змістовні назви підрозділів і дозволяє оперативно працювати з показниками. Наприклад, розділ «Short-term business statistics» має підрозділ «Industry», у якому зручно можна обрати необхідний підрозділ серед переліку: Production in industry, Turnover in industry, Producer prices in industry, Labour input in industry. Однак ця система статистичних даних не дозволяє самостійно відбирати показники для формування аналітичних таблиць (діаграм) з метою їх порівняння. Таблиці показників і їх даних є стаціонарними та незмінними. Вона призначена для дослідження розвитку тільки країн Європейського Союзу, статистичні дані України відсутні.

Статистична база The World bank DataBank³ є найбільш популярною серед вітчизняних користувачів, оскільки містить статистичні дані України. Вона має значну кількість переваг у використанні порівняно з іншими статистичними системами: можливість формування користувачем таблиць і графіків відповідно до потреб (вибір показників, країн, періоду та ін.), зручна форма збереження результатів запитів у різноманітних форматах.

Важливою є можливість збереження результатів у програмі Excel для подальшої роботи з даними. Представлена система

² Eurostat Data Browser (2022). URL: https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes?lang=en&subthme=shorties&display=list&sort=category&extractionId=TEINA021

³ The World bank DataBank (2022). URL: <https://databank.worldbank.org/source/world-development-indicators>

уможливило отримання значної кількості статистичних показників розвитку економіки України та інших країн. Стандарт опису показників (структура статистичних показників) є змістовним і зручним, що дозволяє оперативно знайти необхідний показник у базі за рахунок використання інформативного дерева статистики. Основним обмеженням цієї системи є відсутність значної кількості показників, які характеризують розвиток промисловості України та інших зарубіжних країн, що потребує додаткового пошуку і використання статистичних даних у Державній службі статистики України та інших джерелах.

Дослідження статистичної бази Державної служби статистики України¹ свідчить про наявність значної кількості обмежень у її використанні: інформація розмі-

щується на сайті у вигляді окремих таблиць Excel, що не дає можливості користувачам самостійно обирати для дослідження певні показники та формувати на їх основі таблиці та графіки відповідно до потреб.

Опис показників значною мірою відрізняється від світових стандартів, що обмежує можливості зручного, автоматизованого пошуку показників і даних та дозволяє знаходити їх тільки у відповідних внутрішньо розміщених таблицях, а не у статистичній системі за назвою (ключовими словами), як це демонструє провідний іноземний досвід.

Аналіз функціоналу перелічених автоматизованих систем (сайтів) статистичних баз даних дозволив визначити їх відмітні риси, особливості, переваги і недоліки (табл. 1).

Таблиця 1 – Порівняльний аналіз функціоналу статистичних систем

Статистична база даних	Переваги	Обмеження
1	2	3
1. OECD.Stat ²	Зручність дерева вибору показників; інформативність підрозділів; можливість вибору в таблиці необхідного показника, країни, одиниці вимірювання	Не включено статистичні дані України. Можливість перегляду статистичних даних тільки за одним показником або країною, що ускладнює формування порівняльних таблиць. Показники не структуровані за узагальненими групами, а представлені великим переліком
2. Eurostat Data Browser ³	Має досить наочну структуру та змістовні назви підрозділів, дозволяє оперативно працювати з показниками	Статистичні дані України відсутні. Не дозволяє самостійно відбирати показники для формування аналітичних таблиць (діаграм) з метою їх порівняння. Таблиці показників і їх даних є стаціонарними та незмінними
3. The World bank DataBank ⁴	Можливість формування користувачем таблиць і графіків відповідно до потреб (вибір показників, країн, періоду та ін.), зручна форма збереження результатів запитів у різноманітних форматах. Наявність даних статистики України. Інформативне дерево статистики	Відсутність значної кількості показників, які характеризують розвиток промисловості України та інших країн, що потребує додаткового пошуку і використання статистичних даних у Державній службі статистики України та інших джерелах

¹ Державна служба статистики України (2022). Статистична інформація. URL: <https://ukrstat.gov.ua/>

² OECD.Stat (2022). URL: <https://stats.oecd.org/>

³ Eurostat Data Browser (2022). URL: https://ec.europa.eu/eurostat/databrowser/explore/all/all_themes?lang=en&subthme=shorties&display=list&sort=category&extractionId=TEINA021

⁴ The World bank DataBank (2022). URL: <https://databank.worldbank.org/source/world-development-indicators>

1	2	3
4. Статистична інформація Державної служби статистики України ¹	Наявність статистики України, у тому числі деталізованих статистичних даних з промисловості та інших видів економічної діяльності і галузей.	Статистичні дані розміщені в окремих файлах, що не дає можливості користувачам обирати показники та формувати на їх основі таблиці відповідно до потреб. Опис показників на сайті значною мірою відрізняється від світових стандартів, що обмежує можливості зручного їх пошуку

Джерело: складено авторами.

Таким чином, у більшості існуючих статистичних систем не включено статистичні дані України, а там, де вони є, не представлено деталізацію даних за видами економічної діяльності та галузями, особливо по промисловості. До того ж характерними є технічні труднощі в роботі з даними для користувачів. Ці обставини важливо враховувати при побудові спеціалізованої бази статистичних даних про промисловість.

Порядок формування і технологія наповнення спеціалізованої бази даних про промисловість

Призначенням заявленої інформаційної бази є її використання для відстеження попередніх і поточних тенденцій зміни статистичних даних. Також статистична база даних може забезпечувати автоматизовану смарт-систему промислового прогнозування і форсайтингу через окремий модуль надання даних для економіко-математичного моделювання розвитку промисловості країни.

Для побудови такої бази необхідним є виконання відповідних вимог, а саме:

наявність статистичних даних України та інших країн у єдиній базі;

деталізоване подання показників по Україні за ключовими видами економічної діяльності та галузями, у тому числі з промисловості;

наявність інформативної структури статистичних показників (дерева статистики), розподіленої на розділи, підрозділи тощо;

можливість зручного пошуку статистичних показників за назвою, ключовими словами;

довільний відбір показників і подання їх у вигляді таблиці та/або діаграми (можливість редагування таблиці/діаграми за необхідності);

використання фільтрів для відбору і перегляду показників і їх значень за країнами, секторами економіки, регіонами в межах країни, періодами, одиницями вимірювання тощо;

можливість експорту відібраних наборів статистичних показників у зручних форматах для подальшого використання.

Для формування бази даних проаналізовано показники різних статистичних систем із метою визначення її структури. Показники розподілено за секторами економіки (сектор виробництва сировини, переробний сектор, макростатистика, інші сектори). Зміст секторів розкрито за видами економічної діяльності. Слід зауважити, що основну статистичної бази становлять ключові види економічної діяльності², які забезпечували переважну частину доданої вартості економіки України:

1) сільське, лісове та рибне господарство;

¹ Державна служба статистики України (2022). Статистична інформація. URL: <https://ukrstat.gov.ua/>

² Класифікація видів економічної діяльності (КВЕД-2010) (2022). URL: http://kved.ukrstat.gov.ua/KVED2010/kv10_i.html

- 2) добувна промисловість і розроблення кар'єрів;
- 3) переробна промисловість;
- 4) постачання електроенергії, газу, пари та кондиційованого повітря;
- 5) водопостачання; каналізація, поводження з відходами;
- 6) промисловість усього;
- 7) будівництво.

Кожен визначений вид економічної діяльності розкрито за групами галузей та галузями, яких налічується 48.

Загалом розподіл показників виконано за такими характеристиками: код, сектор, вид економічної діяльності, галузь, одиниці вимірювання, період, база, джерело (табл. 2).

Таблиця 2 – Зміст показників

Англійською мовою	Українською мовою	Зміст
Indicator	Показник	Сформовано на основі назв показників досліджених статистичних баз
Unit	Одиниця вимірювання	Одиниці вимірювання для показників (натуральні, вартісні, відносні)
Branch	Галузь	Галузі за ключовими ВЕД
Activity	Вид економічної діяльності (ВЕД)	Ключові ВЕД
Industry	Ознака промисловості	Позначка про відношення показника до промисловості
Sector	Сектор	Сектори відповідно до структури економіки
Country	Країна	130 країн, відібраних за критеріями: чисельність населення > 1 млрд осіб; ВВП на душу населення > 1 тис. пост. дол. США
Base	База даних	Показники розподілено за такими групами даних відповідно до призначення: статистика, статистика довгострокової моделі, статистика короткострокової моделі, статистика моделі людського капіталу, статистика еко моделі
Source	Джерело даних	Бази даних, із яких використовувались показники: Світовий банк, Держкомстат України, Євростат, НБУ, ФАО, МВФ, TCdata360, розрахункові дані (розраховані авторами)
Period	Період	Календарний період (рік, день, місяць, квартал, півріччя)
Data	Значення показника	Значення показників, які містяться у джерелах даних
Date	Дата	Дата, за якою вносилося значення показника

Джерело: розроблено авторами.

Усі вибрані для бази даних показники з їх значеннями відповідно до періоду та разом із визначеними характеристиками згруповано в одну таблицю, яка має

назву «Grossbuch» та являє собою базу статистичних даних. Структуру запропонованої статистичної бази наведено на рис. 1.

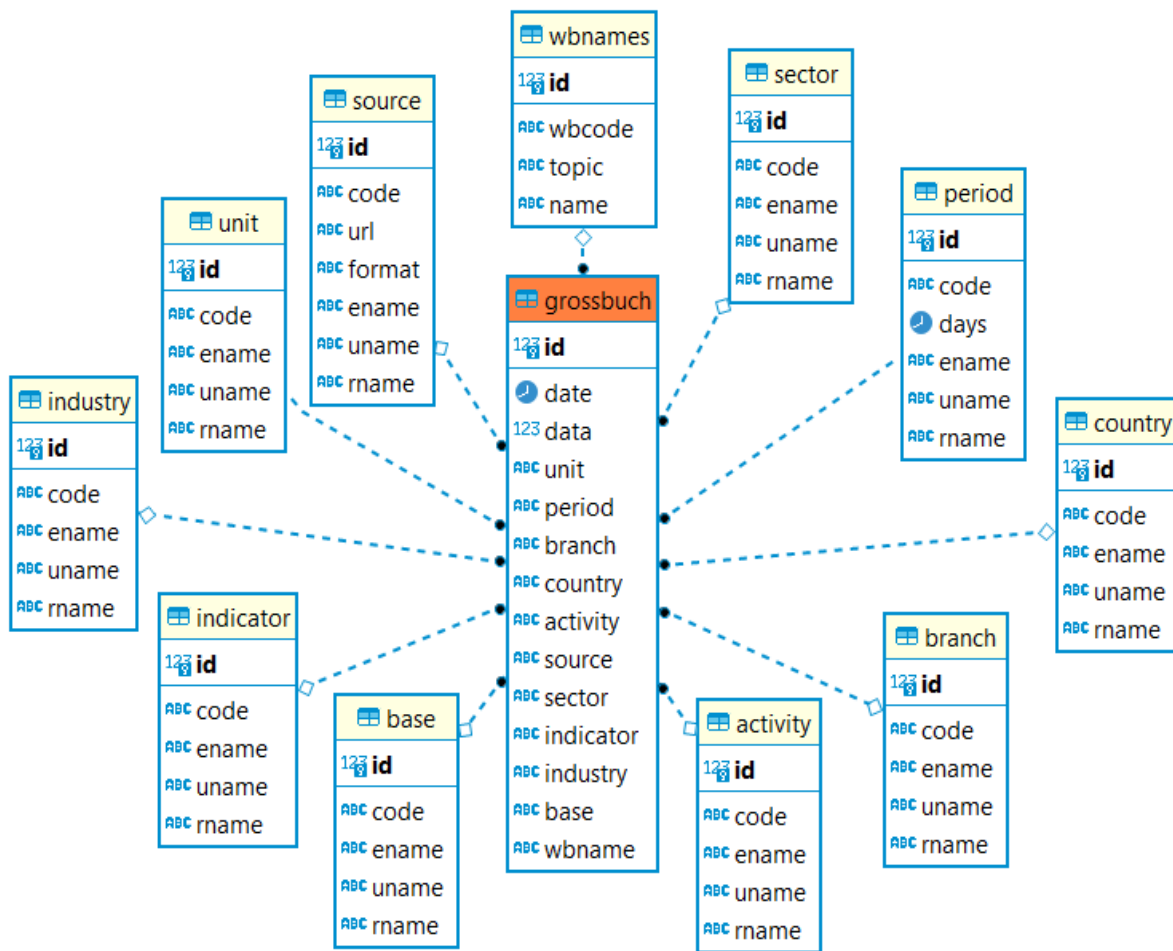


Рисунок 1 – Структура статистичної бази даних про промисловість

Джерело: розроблено авторами.

Базу статистичних даних розвитку промисловості сформовано з використанням: груп показників; статистичних баз, із яких зібрано показники і дані; програмних засобів для роботи з даними, взаємозв'язок яких представлено в узагальненій схемі інформаційного забезпечення статистичної системи (рис. 2). Схему побудовано на основі стандарту IDEF0¹.

У результаті формування бази статистичних даних можна отримати систему показників розвитку промисловості та систему показників за ключовими видами економічної діяльності.

¹ IDEF0 Function modeling method. IDEF Integrated DEFinition methods. URL: https://www.edef.com/idefo-function_modeling_method/

Формування і наповнення статистичної бази даних доцільно здійснювати за такими етапами:

1. *Формування переліку показників.* Цей етап має особливості реалізації для різних статистичних баз. Так, у процесі роботи з базою даних Світового банку із загального переліку вибрано необхідні показники і завантажено їх у форматі готової таблиці Excel. Із сайту Держкомстату України формування переліку показників, їх вибір і внесення значень у загальну таблицю виконано в ручному режимі, оскільки статистичні дані на цьому сайті наведено окремими файлами у форматі Excel, що не дозволяє автоматично формувати перелік разом із значеннями. Таким чином,

сформувалися дві таблиці показників: Світового банку (відібрано тільки перелік показників, оскільки база даних автоматизована і є можливість завантаження значень показників без необхідності їх трансфор-

мації) і Держкомстату України (відібрано перелік показників, трансформовано таблиці Excel з їх значеннями для можливості подальшої роботи з ними).

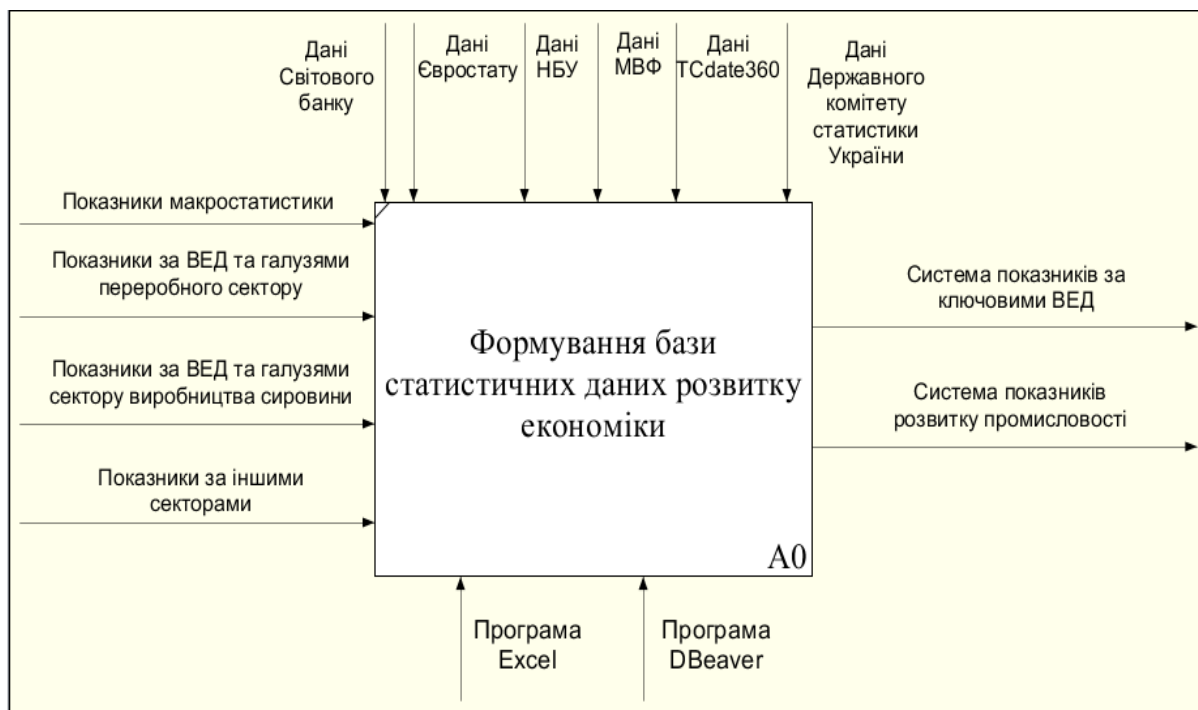


Рисунок 2 – Узагальнена схема інформаційного забезпечення статистичної системи
Джерело: розроблено авторами.

2. *Визначення характеристик кожного з показників* (код, сектор, вид економічної діяльності, галузь, одиниці вимірювання, період, база, джерело, країна), які наведено в табл. 2. В окремій таблиці здійснено загальний збір всіх показників, відібраних для роботи, та розподілено їх за характеристиками в ручному режимі. Для кожного показника визначено його код, обрано сектор, вид економічної діяльності, галузь та інші характеристики.

3. *Створення й оновлення довідників за всіма показниками та їх характеристиками* здійснено шляхом внесення назв показників і їх характеристик українською, англійською та російською мовами, а також присвоєння їм кодів для реєстрації в базі даних у ручному режимі. За новими

показниками інформація в довідниках має оновлюватися.

4. *Створення шаблонів таблиць.* Із метою наповнення даними автоматизованої системи створено шаблони таблиць міграції даних (для показників Світового банку та Держкомстату України). Заповнення шаблонів таблиць даними для імпорту відбувалося автоматично в модулі міграції даних, що працює у напівавтоматичному режимі.

Заповнені таблиці міграції даних можуть передаватися фахівцям (програмістам) для обробки і формування окремого статистичного модуля смарт-системи промислового прогнозування і форсайтингу.

Користувачі при роботі з автоматизованою статистичною базою даних розвитку промисловості мають можливість:

доступу до статистичних даних України та інших зарубіжних країн у єдиній системі;

використання показників за ключовими видами економічної діяльності та галузями по Україні, у тому числі по промисловості;

зручного пошуку статистичних показників за назвою, ключовими словами завдяки наявності інформативної структури даних;

експорту відібраних наборів статистичних показників у зручних форматах для подальшого використання, зокрема MS Excel.

Одним із результатів роботи є формування спеціалізованого програмного забезпечення для автоматизованого збору та поширеного доступу до глобальних статистичних даних про промисловість. Як зазначено вище, автоматизована система, побудована на основі цього програмного

забезпечення, передбачає автоматизований збір інформації за багатьма країнами світу та з використанням будь-якої кількості джерел, доступних як клієнт-серверні ресурси в інтернеті. Різноманітна подача інформації, формат її надання та періодичність оновлення з боку різних джерел не є проблемою, тому що система використовує сучасні алгоритми інтеграції даних. Для зберігання інформації використано сучасну реляційну базу даних, яка дозволяє вирішувати питання оптимізації для роботи з найбільш потужними, але ще не великими даними, уможливорює горизонтальне та вертикальне масштабування, у тому числі в системі баз даних із відкритим вихідним кодом PostgreSQL.

Крім того, просунутим користувачам системи – дослідникам надається збільшений спектр можливостей для конструювання критерію пошуку та формату отримання даних (рис. 3).

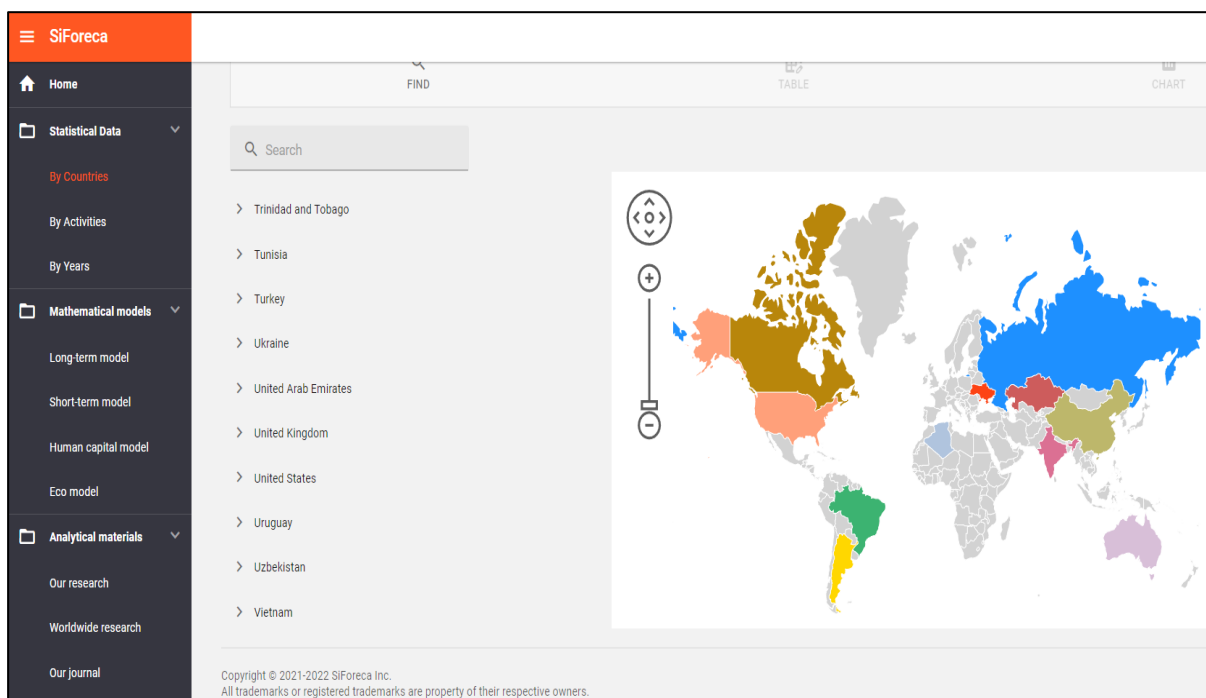


Рисунок 3 – Елемент інтерфейсу користувача статистичної бази даних про промисловість
Джерело: розроблено авторами.

Система побудована як клієнт-серверна технологія та є доступною для без-

шовної інтеграції з аналогічними системами як джерело і сховище інформації, а та-

кож може служити частиною для більш масштабної системи моделювання та прогнозування розвитку промисловості з використанням інструментів економічної аналітики, математичного моделювання тощо. Основу інформаційної системи становлять принципи інтеграції малих, але широких даних (small but wide data¹). У процесі побудови системи вирішено також питання організації різних режимів доступу користувачів та інформаційної безпеки.

Усе це в комплексі дозволяє покращити інформаційне забезпечення аналітичної роботи та процесів прийняття рішень у промисловості та підняти їх на новий рівень.

Висновки. Проблеми інформаційного забезпечення розвитку вітчизняної промисловості є надзвичайно актуальними і важливими. Але на сьогодні в Україні не існує спеціалізованої бази даних про промисловість. Це ускладнює забезпечення необхідними статистичними даними процесів аналітичного дослідження національної індустрії та інших ключових видів економічної діяльності в Україні, а також співставлення їх із даними зарубіжних країн для порівняння.

Для вирішення цієї проблеми розроблено науково-методичний підхід до формування спеціалізованої бази та автоматизованої системи інтеграції даних про промисловість. Її призначення – відстеження попередніх і поточних тенденцій зміни статистичних даних, забезпечення необхідною інформацією модуля економіко-математичного моделювання розвитку промисловості країни, надання користувачам системи можливостей зручного користування наявними статистичними даними для проведення аналітичних досліджень.

Особливість запропонованого підходу полягає в тому, що він передбачає фор-

мування автоматизованої бази даних з орієнтацією на промисловість, яка детально розкриває дані за ключовими видами економічної діяльності та галузями економіки України. Перевагою статистичної системи є можливість порівняння значень показників за 130 країнами й основними характеристиками (період, сектор, вид економічної діяльності, галузь, одиниця вимірювання), а головним недоліком – поки що обмежений функціонал у використанні критеріїв пошуку показників і редагуванні таблиць (діаграм).

Статистичну систему запропоновано розробляти як реляційну базу даних. Такий підхід до її побудови дозволяє за необхідності оновлювати значення показників в автоматизованому режимі, використовувати автономні дані для моделювання тощо, а також формувати будь-які аналітичні матеріали відповідно до потреб користувачів.

Подальший розвиток запропонованої інформаційної системи передбачає розширення автоматизованої бази даних переліком показників шляхом додаткового заповнення таблиць міграції даними із сайтів Світового банку, Держкомстату України та інших статистичних систем. Оскільки створена база даних розвитку промисловості є важливою складовою смарт-системи промислового прогнозування і форсайтингу, яка на сьогоднішній день перебуває у стадії розроблення, у подальшому вона буде розширюватися функціоналом для користувачів (упровадження фільтрів щодо вибору джерел даних, приналежності даних до певного комплексу моделей: короткострокової моделі розвитку промисловості на основі випереджальних індикаторів; довгострокової моделі розвитку промисловості за основними секторами; довгострокової моделі розвитку людського капіталу промисловості; довгострокової моделі екологізації промисловості).

Перспективи подальших досліджень у розглянутій предметній сфері полягають у вдосконаленні процесів майнінгу з вико-

¹ Tcworld (2022). From big data to small and wide data. URL: <https://www.tcworld.info/news-about-technical-communication/from-big-data-to-small-and-wide-data-1128/>

ристанням неструктурованих даних про промисловість, розвитку системи інтеграції даних для надання користувачам інтегрованого доступу до безлічі їх різномірних джерел, що створюватиме в кожного з них відчуття єдиної однорідної бази даних, розробленої для конкретних потреб.

Література

- Вишневецький В. П., Вієцька О. В., Вієцький О. А. та ін. (2019). *Смарт-промисловість: напрями становлення, проблеми і рішення*: монографія. Київ: ІЕП НАН України. 464 с.
- Вишневецький В. П., Вієцька О. В., Гаркушенко О. М. та ін. (2018). *Смарт-промисловість в епоху цифрової економіки: перспективи, напрями і механізми розвитку*: монографія. Київ: ІЕП НАН України. 192 с.
- Дасив А. Ф., Мадых А. А., Охтень А. А., Турлакова С. С. (2019). *Экономико-математические модели и информационно-коммуникационные технологии развития смарт промышленности*: монографія. Киев: ИЭП НАН Украины. 140 с. (электронное издание).
- Полтавцева М. А., Хабаров А. Р. (2016). Безопасность баз данных: проблемы и перспективы. *Программные продукты и системы*. Т. 3 (115). С. 36-41. DOI: <https://doi.org/10.15827/0236-235X.115.036-041>
- Чесалов Л. Е., Шпильман А. В., Аракчев Д. Б. (2016). О состоянии и перспективах информационного обеспечения отрасли. *Разведка и охрана недр*. № 5. С. 155-157.
- Delplanque J., Etien A., Anquetil N., Auverlot O. (2018). Relational Database Schema Evolution: An Industrial Case Study, *IEEE International Conference on Software Maintenance and Evolution (ICSME)*. P. 635-644. DOI: <https://doi.org/10.1109/ICSME.2018.00073>
- Li C., Chen Y., Shang Y. (2022) A review of industrial big data for decision making in

intelligent manufacturing. *Engineering Science and Technology*, Vol. 29. P. 1-16. DOI: <https://doi.org/10.1016/j.jestch.2021.06.001>

- Yin S., Kaynak O. (2015). Big Data for Modern Industry: Challenges and Trends [Point of View]. *Proceedings of the IEEE*. Vol. 103. № 2. P. 143-146. DOI: <https://doi.org/10.1109/JPROC.2015.2388958>
- Zarte M., Pechmann A., Nunes I. L. (2019). Decision support systems for sustainable manufacturing surrounding the product and production life cycle – A literature review. *Journal of Cleaner Production*. Vol. 219. P. 336-349. DOI: <https://doi.org/10.1016/j.jclepro.2019.02.092>

References

- Vyshnevsky, V. (Ed.) (2018). Smart industry in the era of digital economy: prospects, directions and mechanisms of development: monograph / [V. P. Vyshnevsky, O. V. Vietska, O. M. Harkushenko, S. I. Kniaziev, O. V. Liakh, V. D. Chekina, D. Yu. Chervatskyi]; In V. P. Vyshnevsky (Ed.); NAS of Ukraine, Institut of Industrial Economics. Kyiv, 192 p. [in Ukrainian].
- Dasiv, A. F., Madykh, A. A., Okhten, O. O., & Turlakova, S. S. (2019). *Economic-mathematical models and information-communication technologies of smart industry development*: monograph. Kyiv: Institute of Industrial Economics of NAS of Ukraine. 140 p. (electronic edition) [in Russian].
- Poltavtseva, M. A., & Khabarov, A. R. (2016). Database security: problems and prospects. *Software products and systems*, 3 (115), pp. 36-41. DOI: <https://doi.org/10.15827/0236-235X.115.036-041> [in Russian].
- Chesalov, L. E., Shpilman, A. V., & Arakcheev, D. B. (2016). On the state and prospects of information support for the industry. *Exploration and protection of subsoil*, 5, pp. 155-157 [in Russian].

- Delplanque, J., Etien, A., Anquetil, N., & Auverlot, O. (2018). Relational Database Schema Evolution: An Industrial Case Study, *IEEE International Conference on Software Maintenance and Evolution (ICSME)*. pp. 635-644. DOI: <https://doi.org/10.1109/ICSME.2018.00073>
- Li, C., Chen, Y., & Shang, Y. (2022). A review of industrial big data for decision making in intelligent manufacturing. *Engineering Science and Technology*, 29, pp. 1-16. DOI: <https://doi.org/10.1016/j.jestch.2021.06.001>
- Yin, S., & Kaynak, O. (2015). Big Data for Modern Industry: Challenges and Trends [Point of View]. *Proceedings of the IEEE*, 103 (2), pp. 143-146. DOI: <https://doi.org/10.1109/JPROC.2015.2388958>
- Zarte, M., Pechmann, A., & Nunes, I. L. (2019). Decision support systems for sustainable manufacturing surrounding the product and production life cycle – A literature review. *Journal of Cleaner Production*, 219, pp. 336-349. DOI: <https://doi.org/10.1016/j.jclepro.2019.02.092>
- Vyshnevsky, V. P., Viietska, O. V., & Viietskyi, O. A. (Ed.) (2019). *Smart industry: directions of development, problems and solutions*: monograph. K. NAS of Ukraine, Institut of Industrial Economics. 464 p. [in Ukrainian].

Алла Федоровна Дасив,

канд. экон. наук

Институт экономики промышленности НАН Украины
ул. Марии Капнист, 2, г. Киев, 03057, Украина

E-mail: alladasiv@gmail.com

<https://orcid.org/0000-0001-5431-701X>;

Елена Анатольевна Руссиян,

канд. экон. наук, старший научный сотрудник

Институт экономики промышленности НАН Украины
ул. Марии Капнист, 2, г. Киев, 03057, Украина

E-mail: elena.russiyana@gmail.com

<https://orcid.org/0000-0002-1246-9952>;

Денис Владимирович Липницкий,

канд. экон. наук, ведущий инженер по разработке ПО

Vodafone Germany GmbH, г. Дюссельдорф, 40001, Германия

E-mail: den.lipnitsky@gmail.com

<https://orcid.org/0000-0002-4616-7936>

ФОРМИРОВАНИЕ БАЗЫ И АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ИНТЕГРАЦИИ ДАННЫХ О ПРОМЫШЛЕННОСТИ

В статье обоснована целесообразность формирования специализированной базы данных о промышленности и системы автоматизированного их сбора и использования. Эта система позволяет обеспечить информационные потребности аналитиков и учёных по проблемам развития промышленности. Она предусматривает автоматизированный сбор информации по многим странам мира с использованием любого количества источников, доступных в качестве клиент-серверных ресурсов в интернете. Применение современных алгоритмов интеграции данных обеспечивает разнообразную подачу информации, форматы ее предоставления и периодичность обновления. Кроме того, продвинутым пользователям системы предоставляется широкий круг возможностей для создания критерия поиска и формата получения данных. Система построена как клиент-серверная технология и до-

ступна для бесшовной интеграции с аналогичными системами как источник и хранилище информации.

Для построения системы проведен анализ существующих в мире и популярных в использовании статистических баз данных, определены их особенности, преимущества и недостатки. Установлено, что в большинстве существующих статистических систем не представлены статистические данные по Украине, а там, где они есть, отсутствует детализация данных по видам экономической деятельности и отраслям, особенно по промышленности. К тому же имеются определенные технические сложности в работе с данными для пользователей, персонализация доступа почти отсутствует. Достаточно ограничены возможности интеграции существующих баз данных с системами, использующими информацию для моделирования и прогнозирования как в режиме запроса, так и особенно в режиме реального времени.

Сформирована структура статистической базы данных о развитии промышленности, отобраны показатели для ее наполнения. Показатели распределены по секторам экономики (сектор производства сырья, перерабатывающий сектор, макростатистика, другие секторы). Содержание секторов раскрыто по ключевым видам экономической деятельности в соответствии с классификацией видов экономической деятельности 2010 года (КВЭД-2010), которые обеспечивают преобладающую часть создания добавленной стоимости экономики Украины. Каждый определенный вид экономической деятельности раскрыт по группам отраслей и отраслям.

Кроме того, база статистических данных сформирована на основе присвоения каждому из показателей других характеристик, необходимых для автоматизации: код, единицы измерения, период, база (распределение показателей по отдельным группам для возможности моделирования), источник (статистические базы и источники, с которых осуществлялся сбор показателей и данных).

Для хранения информации использована современная реляционная база данных, которая позволяет решать вопросы оптимизации для работы с наиболее мощными, но еще не большими данными, учитывает особенности широких данных, дает возможность горизонтального и вертикального масштабирования, в том числе в системе баз данных с открытым исходным кодом PostgreSQL.

Создана методика формирования и технология наполнения статистической базы данных автоматически из большого количества источников, режим доступа к которым задается параметрами конфигурации системы. Разработана технология миграции данных, доступных в гибких форматах, в том числе текстовом, в частности Excel.

Использование разработанной автоматизированной информационной системы интеграции промышленной статистики позволяет выбирать необходимые показатели для анализа экономических процессов в промышленности, использовать собранные в едином информационном пространстве статистические данные по 130 странам мира, в том числе Украины, для проведения научных исследований, построения экономико-математических моделей и принятия управленческих решений.

Ключевые слова: промышленность, экономические характеристики, статистика, база данных, автоматизированная система интеграции данных, информационная разработка, экономическая аналитика.

JEL: L86, L6, L52, L16, F47, E27, C82, C81

Alla F. Dasiv,

PhD in Economics

Institute of Industrial Economics of NAS of Ukraine

2 Maria Kapnist Street, Kyiv, 03057, Ukraine

E-mail: alladasiv@gmail.com

<https://orcid.org/0000-0001-5431-701X>;

Olena A. Russiian,

PhD in Economics, Leading Researcher

Institute of Industrial Economics of NAS of Ukraine

2 Maria Kapnist Street, Kyiv, 03057, Ukraine

E-mail: elena.russiyan@gmail.com

<https://orcid.org/0000-0002-1246-9952>;

Denys V. Lypnytskyi,

PhD in Economics, Lead Software Development Engineer

Vodafone Germany GmbH, Düsseldorf, 40001, Germany

E-mail: den.lipnitsky@gmail.com

<https://orcid.org/0000-0002-4616-7936>

FORMATION OF THE BASE AND THE AUTOMATED SYSTEM OF THE DATA INTEGRATION ABOUT THE INDUSTRY

The expediency of forming the specialized database about industry and the system of their automated collection and usage were justified in this paper. This system allows to provide the information needs of the analysts and the scientists about the problems of industrial development. It provides the automated collection of the information for many countries of the world using any number of sources available as client-server resources on the Internet. The usage of modern data integration algorithms provides a diverse presentation of information, formats for its provision and frequency of updates. In addition, the advanced users of the system are provided with a wide range of options for creating search criteria and data acquisition format. The system is built as a client-server technology and is available for seamless integration with similar systems as a source and storage of information.

In order to build a system, the analysis of existing in the world and popular in use statistical databases was carried out, their features, advantages and disadvantages were defined. It has been established that most of the existing statistical systems do not provide statistical data for Ukraine, and where they are, there is no detail data by type of economic activity and industry, and especially in industry. In addition, there are certain technical difficulties in working with data for users, the personalization of access is almost non-existent. The possibilities of the integrating existing databases with the systems that use information for modeling and forecasting both in query mode and especially in real time are rather limited.

The structure of the statistical database about the development of industry has been formed, the indicators have been selected for its filling. The indicators were distributed by the sectors of the economy (the raw material production sector, the processing sector, the macrostatistics, other sectors). The sectors were disclosed by the key types of economic activity in accordance with the classification of economic activities in 2010 (CEA-2010), providing the predominant part of the creation of the added value of the economy of Ukraine. Each specific type of economic activity is disclosed by groups of industries and industries. In addition, the formation of the statistical database was carried out on the basis of assigning to each of the indicators other characteristics necessary for the automation: code, units of measurement, period, base (distribution of the indicators into the separate groups for the possibility of modeling), source (the statistical bases and sources

from which the indicators and the data were collected). The modern relational database was used to store information, which allows solving optimization issues for working with the most powerful, but not yet large data, taking into account the features of wide data, allows horizontal and vertical scaling, including in the PostgreSQL open source database system.

The methodology for the formation and the technology for filling the statistical database automatically from a large number of sources, the access mode to which is set by the system configuration parameters has been created. The technology has been developed for data migrations available in flexible formats, including text, in particular Excel.

The implementation of the developed automated information system integration of industrial statistics allows you to select the necessary indicators for the analysis of economic processes in industry, use the statistical data collected in a single information space for 130 countries of the world (including Ukraine) for conducting scientific researches, building economic and mathematical models and making managerial decisions.

Keywords: industry, economic characteristics, statistics, database, automated data integration system, information development, economic analytics.

JEL: L86, L6, L52, L16, F47, E27, C82, C81

Формат цитування:

Дасів А. Ф., Руссиян О. А., Липницький Д. В. (2022). Формування бази та автоматизованої системи інтеграції даних про промисловість. *Економіка промисловості*. № 3 (99). С. 57-71. DOI: <http://doi.org/10.15407/econindustry2022.03.057>

Dasiv, A. F., Russian, O. A., & Lypnytskyi, D. V. (2022). Formation of the base and the automated system of the data integration about the industry. *Econ. promisl.*, 3 (99), pp. 57-71. DOI: <http://doi.org/10.15407/econindustry2022.03.057>

Надійшла до редакції 03.08.2022 р.