

КІЛЬКІСНІ МЕТОДИ В СОЦІАЛЬНО-ДЕМОГРАФІЧНИХ ДОСЛІДЖЕННЯХ

QUANTITATIVE METHODS IN SOCIO-DEMOGRAPHIC RESEARCH

DSE | DEMOGRAPHY
AND
SOCIAL ECONOMY

Cite: Rozbytskyi, Mykhailo (2024). Vykorystannia velykykh danykh dlia informatsiinoho zabezpechennia rynku pratsi [Usage of Big Data for Information Support of the Labor Market]. *Demohrafiia ta sotsialna ekonomika - Demography and Social Economy*, 1 (55), 133-147. <https://doi.org/10.15407/dse2024.01.133>



<https://doi.org/10.15407/dse2024.01.133>

УДК: [004.738.5: 331.5]-043.86

JEL Classification: J21, C55, O33

РОЗБИЦЬКИЙ МИХАЙЛО, голов. економіст

Інститут демографії та проблем якості життя НАН України

01032, Україна, м. Київ, бул. Т. Шевченка, 60

E-mail: rozbytskyi.m@gmail.com

ORCID: 0009-0007-4102-0912

ВИКОРИСТАННЯ ВЕЛИКИХ ДАНИХ ДЛЯ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ РИНКУ ПРАЦІ

Проблеми збору актуальних даних щодо ринку праці в умовах глобальних змін та геополітичних коливань потребують нових підходів та альтернативних методів збору і аналізу інформації. Це підкреслює актуальність досліджень, спрямованих на розвиток та адаптацію методів формування інформаційного забезпечення ринку праці до сучасних викликів. Перспективним підходом є використання великих даних (Big Data) для оцінки ринку праці, що включає збір інформації з таких джерел як онлайн-портали з пошуку роботи та вакансій. Підхід дає змогу глибше аналізувати ринкові тенденції та забезпечувати більш точну й своєчасну оцінку потреб і можливостей на ринку праці. Метою даної статті є висвітлення підходів до формування інформаційного забезпечення ринку праці на основі використання великих даних, зокрема онлайн даних з сайтів вакансій. Розглянуто особливості використання великих даних для аналізу ринку праці, на основі бази даних, яка вміщує понад чотири мільйони вакансій, розміщених на українських порталах пошуку роботи за останні п'ять років, наданих Європейським фондом освіти — European Training Foundation (ETF). Здійснено оцінку ефективності цих підходів для сприяння пошуку роботи для усіх зацікавлених осіб, зокрема шляхом надання інформації про динаміку попиту та пропозиції на ринку праці на основі даних, отриманих з порталів. Проаналізовано можливість та обмеження використання великих даних, включаючи аналіз їх впливу на розвиток політики

© Видавець ВД «Академперіодика» НАН України, 2024. Стаття опублікована за умовами відкритого доступу за ліцензією CC BY-NC-ND (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

© Publisher PH «Akademperiodyka» of the NAS of Ukraine, 2024. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

зайнятості та планування ринку праці. Окреслено потенційні переваги великих даних у забезпеченні глибшого і більш точного аналізу ринкових умов, а також розглянуто технічні аспекти та виклики, пов'язані з їх обробкою та інтерпретацією. У статті проаналізовано методологічні підходи до збору даних та аналітичного оброблення у контексті прискорених трансформацій, волатильності та обмеженого доступу до традиційних інформаційних ресурсів. Наукова новизна статті полягає в обґрунтуванні можливості та доцільності використання відкритих даних з онлайн-порталів вакансій для інформаційного забезпечення ринку праці в сучасних умовах. У дослідженні застосовано методи аналізу, синтезу та узагальнення для визначення основних сучасних проблем інформаційного забезпечення ринку праці України. Здійснено оцінку ефективності застосування методів збору даних на основі методів веб-скрейпінгу та парсингу, а також застосування інтегрованої платформи Snowflake з метою виявлення ключових тенденцій та патернів на ринку праці. У висновках узагальнено основні положення та обґрунтовано напрями подальших досліджень, указано на значення великих даних для розвитку стратегій зайнятості та оптимізації ринку праці.

Ключові слова: великі дані, ринок праці, веб-скрейпінг, парсинг, зайнятість, тенденції, вакансії, попит.

Постановка проблеми та актуальність. Від початку повномасштабної війни в Україні, відповідно до Закону України «Про захист інтересів суб'єктів подання звітності та інших документів у період дії воєнного стану або стану війни» від 03 березня 2022 р., використання традиційних методів збору статистичних даних щодо ситуації на ринку праці, що застосовувалися Державною службою статистики України (далі — Держстат), як-от вибіркові обстеження домогосподарств, статистична звітність, унеможливились. Через воєнні дії та руйнування інфраструктури, доступ до інших офіційних даних, зокрема адміністративних даних, став більш обмеженим. У цьому контексті з'явилася нагальна потреба у використанні альтернативних джерел даних. Зокрема, це передбачає доцільність застосування технологій використання великих даних з метою оцінки показників ринку праці. Термін «великі дані» був вперше сформульований в 1990-х рр. Джоном Маші (John Mashey), віцепрезидентом *Silicon Graphics*. У своїй презентації на щорічній технічній конференції *USENIX* 1999 р. в Каліфорнії [1] він описував масивні набори даних, які не можуть бути ефективно оброблені звичайними базами даних або методами обробки інформації. Використання великих даних у сучасному контексті дозволяє аналізувати інформацію з різних джерел, включаючи соціальні медіа, супутникові знімки, мобільні дані та інші [2]. Відомий американський економіст Радж Четті (Raj Chetty) використовував великі дані для аналізу соціальної мобільності та нерівності доходів. Однією з його ключових робіт є дослідження «The Impacts of Neighborhoods on Intergenerational Mobility» [3], де проаналізовано вплив місця проживання на економічні перспективи дітей.

Необхідно зазначити, що в контексті сучасної ери цифрових технологій обсяги доступної інформації зростають експоненціально, що сприяє розширенню можливостей для проведення всебічних наукових досліджень.

Використання передових методик обробки даних стає критичним для ефективного аналізу великомасштабних даних та виведення значущих висновків. Це, зі свого боку, накладає певні вимоги до забезпечення високого рівня конфіденційності оброблюваних даних та дотримання етичних норм у їх обігу. Сучасні дослідники та аналітики зіштовхуються з необхідністю постійного удосконалення технічних компетенцій та зростаючою соціальною відповідальністю у контексті роботи з цифровими даними, що має вирішальне значення для розвитку наукового поступу. Враховуючи нагальну потребу в актуальній інформації, цей напрямок досліджень стає особливо важливим для розвитку стратегій управління ринком праці та формування ефективної економічної політики.

Аналіз останніх досліджень і публікацій. В рамках авторського дослідження було проведено ретельний аналіз масиву літератури, що стосується предмета дослідження та аналогічних досліджень у дотичних сферах. Прикладними дослідженнями займається *ETF*, аналітика на основі їх бази великих даних набуває все більшого поширення в наукових роботах останніх років. На мій погляд, найзначущими є висновки, сформульовані в роботах: В. Саріогло (V. Sarioglo), О. Цимбала (O. Tsymbal) [4], А. Ваккаріно (A. Vaccarino) [5, 6], М. Боузала (M. Boughzala) [7], С. Дусі (S. Dusi), К. Граціані (C. Graziani), М. Меццанцаніка (M. Mezzanzanica) та М. Пелуччі (M. Pelucchi) [8], Ф. Джамбона (F. Giambona), А. Халаві (A. Khalawi), Л. Буцціголі (L. Buzzigoli), Л. Грассіні (L. Grassini), К. Мартеллі (C. Martelli) [9].

Мета дослідження — розробка підходів до інформаційного забезпечення ринку праці на основі використання великих даних з урахуванням наявних обмежень. Це передбачає аналіз обмежень та створення детальної інформації про ключові аспекти ринку праці, як-от попит, пропозиція, зайнятість та тенденції тощо.

Наукова новизна статті полягає в обґрунтуванні доцільності використання відкритих даних з онлайн-порталів вакансій для інформаційного забезпечення аналізу ринку праці в сучасних умовах. Важливим аспектом дослідження є оцінка ефективності цих підходів у контексті сприяння пошуку роботи для шукачів, зокрема через аналіз динаміки попиту та пропозиції на ринку праці на основі даних, отриманих з цих порталів. У роботі використовувалась структурована база даних щодо вакансій, яка була зібрана організацією *ETF*. Дані були сформовані на базі онлайн-порталів вакансій для України впродовж останніх п'яти років.

Методологічна основа дослідження полягає в застосуванні комбінації теоретичних та емпіричних методів. Теоретичні методи включають аналіз, синтез та узагальнення, на основі яких визначено основні сучасні проблеми інформаційного забезпечення ринку праці України та роль великих даних. Емпіричні методи, такі як веб-скрейпінг та парсинг, а також застосування платформи *Snowflake* забезпечили можливість виявлення ключових

тенденцій та патернів на ринку праці, детального висвітлення практичних підходів до використання великих даних для заповнення прогалів у інформаційному забезпеченні ринку праці. Табличний і графічний методи застосовано для візуалізації результатів аналізу.

Виклад основного матеріалу. У сучасному світі, що характеризується стрімкими змінами та цифровою трансформацією, роль інформації як ключового ресурсу стає дедалі значущішою. Великі дані становлять собою великі об'єми інформації, які перевищують здатність звичайних програмних інструментів до їх обробки та аналізу, набуваючи важливості у сферах, що потребують глибокого дослідження, точності та оперативності рішень [10]. Це особливо стосується інформаційного забезпечення ринку праці, де великі дані уможливають виявлення тенденцій у попиті на певні професії та навички, відіграючи ключову роль у розвитку політик зайнятості та сприянні економічному зростанню. В умовах глобалізації та швидких технологічних змін дослідження великих даних стає необхідним для прогнозування майбутніх тенденцій ринку праці.

Наукове співтовариство демонструє щораз більший інтерес до можливостей, які надають великі дані у контексті дослідження ринку праці. Використання великих даних зумовлено їхнім потенціалом у виявленні трендів та закономірностей, які традиційні методи можуть не відслідковувати або ігнорувати. Використання великих даних дає змогу дослідникам змінити підходи до оцінки стану ринку праці, зокрема до формування релевантного інформаційного забезпечення, з метою оперативної оцінки попиту на робочу силу, визначення впливу технологічних інновацій, глобалізації економіки, демографічних змін, а також оцінки впливу соціально-економічних криз, включаючи військові конфлікти.

Широкі перспективи використання великих даних для інформаційного забезпечення ринку праці відкриває використання структурованої бази даних вакансій з онлайн-порталів з пошуку роботи в Україні, сформованої *ETF* перед самим початком війни. Зусилля *ETF* орієнтовані на підтримку країн, які межують з ЄС, у поліпшенні систем професійної освіти та навчання, а також в аналізі потреб у навичках, що є критичним для розвитку ринку праці. Важливість діяльності *ETF* полягає у зміцненні соціальної згуртованості та сприянні сталому економічному розвитку, що теж позитивно впливає на економічні відносини між державами-членами ЄС та їх сусідами.

При формуванні бази даних *ETF* використано метод (технологію) веб-скрейпінгу (*Web Scraping*), що дало можливість об'єднати понад чотири мільйони записів про вакансії в Україні, зібрані за період з 2018 по 2023 рр. зі всіх значущих сайтів вакансій. Таким чином, база даних відображає динаміку ринку праці, що може бути корисним для дослідження змін, зокрема викликаних повномасштабною війною, пандемією *Covid-19* та іншими

економічними і соціальними факторами. Обсяг бази даних перевершує обсяг стандартних статистичних даних, потенційно забезпечуючи достатньо широкі можливості для аналізу тенденцій щодо вакансій, необхідних кваліфікацій та навичок, затребуваних на сучасному ринку праці. Наприклад, детальний аналіз вакансій щодо навичок за класифікацією *ESCO*, яка забезпечує можливість деталізованої оцінки навичок у різних секторах економіки, відображає відповідні тенденції у зміні потреб ринку праці.

До бази *ETF* включено широкий спектр ознак вакансій: загальні ідентифікатори вакансій (*GENERAL_ID*), дати оновлення інформації (*REFRESH_DATE*) та категорії джерел даних (*IDSOURCE_CATEGORY*, *SOURCE_CATEGORY*), індексації дат (*INDEX_DATE*), які охоплюють рік, місяць та день публікації вакансій, а також дати їх включення до бази даних (*GRAB_DATE*) та закінчення терміну дії вакансій (*EXPIRE_DATE*). База даних також містить інформацію про мову оголошення (*LANG*) та рівні кваліфікацій за класифікацією *ESCO*, які поділяються на чотири рівні (*IDESCO_LEVEL_1* до *IDESCO_LEVEL_4*). Додатково фіксується інформація про географічне розташування вакансій, зокрема місто (*IDCITY*, *CITY*), провінцію (*IDPROVINCE*, *PROVINCE*), регіон (*IDREGION*, *REGION*), макрорегіон (*IDMACRO_REGION*, *MACRO_REGION*) та країну (*IDCOUNTRY*, *COUNTRY*).

Щодо характеристик самої роботи: база даних включає тип контракту (*IDCONTRACT*, *CONTRACT*), рівень освіти (*IDEDUCATIONAL_LEVEL*, *EDUCATIONAL_LEVEL*), сектор економіки (*IDSECTOR*, *SECTOR*), макросектор (*IDMACRO_SECTOR*, *MACRO_SECTOR*) та категорію сектору (*IDCATEGORY_SECTOR*, *CATEGORY_SECTOR*). Інформація про заробітну плату (*IDSALARY*, *SALARY*, *SALARYVALUE*), графік роботи (*IDWORKING_HOURS*, *WORKING_HOURS*) та необхідний досвід (*IDEXPERIENCE*, *EXPERIENCE*) дозволяє глибше аналізувати умови праці. Також включені дані про країну джерела (*SOURCECOUNTRY*) та назви компаній (*COMPANYNAME*, *COMPANYRAW*), що надають вакансії. Загалом база налічує 55 змінних.

Використання даних щодо вакансій допомагає не лише ідентифікувати найбільш затребувані професії та ключові навички для успішного працевлаштування, але й визначити актуальні напрями адаптації освітніх програм до сучасних потреб ринку праці. За нинішніх умов існує нагальна потреба в оновленні освітніх програм університетів та коледжів, зосередженні уваги на розвитку навичок, які відповідають потребам сучасного ринку праці. Це також означає, що система професійної освіти та навчання має враховувати швидкість змін у попиті на робочу силу та бути готовими до оперативного впровадження актуалізованих курсів та спеціалізацій, що відповідають новим трендам змін професій та кваліфікацій.

В умовах посилення глобальних викликів, зокрема в Україні, адаптація нових методів аналізу даних стає вирішальною. Традиційні методи збору інформації часто виявляються недостатньо ефективними через обмежений

доступ або ненадійність джерел. У такому контексті методи парсингу (*Parsing*) та веб-скрейпінгу набувають особливої актуальності як альтернативні підходи до збору даних. Точне визначення винахідника цих технологій є складним, оскільки вони розвивалися поступово з появою Інтернету. Однак важливо відзначити, що веб-скрейпінг та парсинг зазнали значного розвитку завдяки роботі багатьох програмістів та інженерів у сфері комп'ютерних наук.

Веб-скрейпінг є процесом автоматизованого збору даних безпосередньо з інтернету. Він включає використання скриптів або програм для автоматичного витягування інформації з вебсторінок. Цей метод особливо ефективний для збору даних з великої кількості вебсайтів. Наприклад, веб-скрейпер може використовувати мову запитів — *XPath*, або спеціальну мову стилю вебсторінок — *CSS* селектори, для визначення конкретних елементів на сторінці, які потрібно витягнути. Приклад *XPath* селектора, з офіційної бібліотеки *XML Path Language* [11], може мати форму:

$$\text{xpath} = '//div[@class="content"]//text()' \quad (1)$$

де: *xpath* — змінна, яка зберігає *XPath* вираз, *//* — символ *XPath* використовується для пошуку елементів у будь-якому місці документа, *div[@class=»content»]* — частина вказує на елемент 'div' з атрибутом 'class', який має значення «content», тобто вираз шукає всі 'div' елементи, чий клас дорівнює «content». *//text()* — функція *XPath*, яка вибирає всі текстові вузли дочірніх елементів вибраного 'div' елемента, це означає, що вираз поверне весь текст, що міститься всередині кожного вибраного 'div' елемента.

Парсинг передбачає використання спеціальних алгоритмів для екстракції та структурування даних з різноманітних неструктурованих та напівструктурованих джерел, у той час як веб-скрейпінг зосереджений на автоматизованому зборі даних із вебсторінок. Наприклад, парсинг *HTML* дає можливість ефективно вилучати інформацію з вебсторінок за допомогою регулярних виразів, описаних в офіційній бібліотеці мови *Python* [12]. Найпростіший вираз може мати вигляд:

$$\text{regex} = r'\langle\text{tag}\rangle(.*)\langle\text{tag}\rangle' \quad (2)$$

де: *regex* — змінна, яка зберігає регулярний вираз, *r* — «сирі рядки» в *Python*, які використовуються для спрощення запису регулярних виразів, оскільки вони не обробляють спеціальні символи як екрановані, *<tag>* — символізує початок *HTML*-тега, *(.*)* — це особлива конструкція в регулярних виразах, яка використовується для знаходження і захоплення будь-якого тексту, що знаходиться між вказаними тегами, *</tag>* — символізує закінчення *HTML*-тега.

Сучасні технології парсингу включають розпізнавання тексту, обробку природної мови та аналіз даних, що уможлиблює виявлення значимих

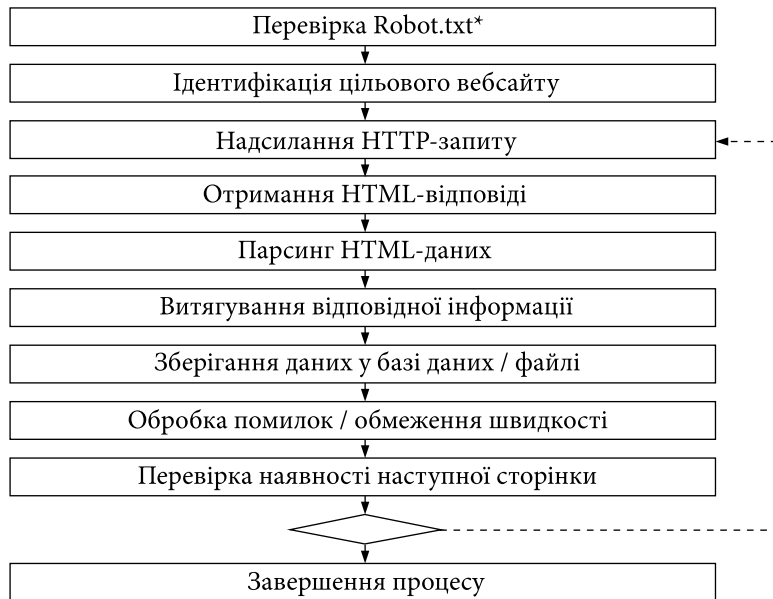


Рис. 1. Процес веб-скрейпінгу

Джерело: авторська розробка на основі детального алгоритму [12].

шаблонів та закономірностей у великих масивах інформації. Основним і базовим процесом веб-скрейпінгу є парсинг обраної сторінки і витягування потрібної інформації (рис. 1).

Діаграму побудовано відповідно до загальноприйнятої практики веб-скрейпінгу. *Robot.txt — це файл веброзробників, що знаходиться в кореневому каталозі сайту і містить інструкції для веброботів, у яких вказується, які частини сайту не можна сканувати або обробляти. Під час використання парсингу та веб-скрейпінгу важливо враховувати, що збір даних відбувається відповідно до законодавства про захист даних та авторські права. Це включає в себе використання надійних методів анонімізації даних та забезпечення, що зібрані дані використовуються виключно у дослідницьких або аналітичних цілях. Крім того, важливо враховувати технічні виклики, пов'язані з парсингом та веб-скрейпінгом: обробку великих обсягів даних, вирішення проблем з сумісністю форматів та забезпечення точності та актуальності інформації.

Для дослідження зібраних даних використовуються технологічні платформи, зокрема, хмарна платформа *Snowflake*. Ця платформа забезпечує потужні інструменти для зберігання, управління та аналізу великих обсягів даних, що дає змогу ефективно використовувати можливості великих даних для аналізу ринку праці. Робота зі *Snowflake* передбачає кілька ключових етапів. Початковим кроком є реєстрація облікового запису та налаштування середовища. Наступним етапом є завантаження даних, яке

```

1 SELECT
2     ESCO_LEVEL_4 AS Profession,
3     YEAR_INDEX_DATE AS Year,
4     COUNT(*) AS JobCount,
5     AVG(SALARYVALUE) AS AverageSalary
6 FROM
7     "LIGHTCAST"."VIEWS_FOR_SHARES"."ETF_UA_EUROPEAN_POSTING"
8 WHERE
9     ESCO_LEVEL_4 IN (
10        'Applications programmers',
11        'Database and network professionals not elsewhere classified',
12        'Database designers and administrators',
13        'Research and development managers',
14        'Software developers',
15        'Web and multimedia developers'
16    )
17    AND SALARYVALUE > 1000 AND SALARYVALUE < 1500000
18 GROUP BY
19     ESCO_LEVEL_4, YEAR_INDEX_DATE
20 ORDER BY
21     ESCO_LEVEL_4, YEAR_INDEX_DATE;

```

Рис. 2. Поле для введення SQL-запитів з прикладом запиту до бази даних
Джерело: скріншот робочого інтерфейсу Snowflake.

може бути здійснене через пакетне завантаження або потокову передачу даних у реальному часі з різноманітних джерел. Організація даних у *Snowflake* включає використання схем, таблиць та переглядів, а також реалізацію заходів управління даними та безпеки.

Запити до даних виконуються за допомогою SQL [13], що дозволяє ефективно витягувати необхідні зв'язки. *Snowflake* надає інструменти для аналізу даних, а також можливість інтеграції зі сторонніми аналітичними інструментами. Однією з ключових можливостей платформи є спільне використання даних, що сприяє безпечному обміну даними з іншими користувачами *Snowflake* або зовнішніми споживачами.

В основі інтерфейсу лежить принцип «що бачиш, те й отримуєш» (*WY-SIWYG*), який забезпечує візуальну консистенцію між тим, як запити відображаються в редакторі та як вони будуть виконані системою. Центральне місце в інтерфейсі займає поле для введення SQL-запитів (рис. 2), оснащене підсвічуванням синтаксису, що робить процес написання коду більш інтуїтивним. Підсвічування синтаксису допомагає уникнути помилок, виділяючи ключові слова, функції та оператори, що забезпечує швидке виявлення та виправлення помилок у запитах.

Результатом запиту є структурована таблиця, яка відображає вибрані дані в організованому форматі для подальшого аналізу та обробки (табл. 1).

Snowflake оптимізує процес обробки SQL-запитів, використовуючи передові алгоритми та кешування результатів. Це забезпечує швидке ви-

конання запитів, навіть коли мова йде про великі обсяги даних. Така ефективність обробки запитів є критично важливою для аналітичних операцій, оскільки вона дозволяє користувачам швидко отримувати зв'язки та виявляти значущі відомості з їхніх даних, підвищуючи загальну продуктивність аналітичних процесів.

Масштабування та оптимізація є важливими аспектами роботи з *Snowflake*, де користувачі можуть регулювати свої обчислювальні ресурси (склади) для адаптації до вимог робочого навантаження. Регулярний моніторинг *Snowflake* дозволяє оптимізувати продуктивність, безпеку та витрати. Важливим елементом є також реалізація стратегій резервного копіювання та відновлення після збоїв для забезпечення доступності та цілісності даних.

Аналітика на основі даних *ETF*, здійснювана через *Snowflake*, дає можливість не лише аналізувати поточний стан ринку праці, але й виявляти тренди та передбачати майбутні зміни. Це важливо для розробки ефективних стратегій у сфері зайнятості та професійного навчання, а також для адаптації до змінюваних ринкових умов. Зокрема, аналіз даних може виявити зростаючі сектори економіки, що потребують додаткової робочої сили, а також сектори, що зазнають спаду. Крім того, *Snowflake* сприяє розумінню змін у вимогах до кваліфікацій та навичок, що є критично важливим для планування освітніх ініціатив та розвитку навчальних програм, орієнтованих на реальні потреби ринку праці.

Таблиця 1. Приклад отриманих даних у сфері інформаційних технологій за останні 4 роки

Професія	Рік	Кількість вакансій	Середня заробітна плата (EUR)
Прикладні програмісти	2020	2389	3495,31
Прикладні програмісти	2021	3064	3619,99
Прикладні програмісти	2022	1862	5433,68
Прикладні програмісти	2023*	1391	6915,59
Менеджери з досліджень та розробок	2020	7687	4530,84
Менеджери з досліджень та розробок	2021	10 206	4973,55
Менеджери з досліджень та розробок	2022	6341	5585,46
Менеджери з досліджень та розробок	2023*	3025	10 168,15
Веб- та мультимедійні розробники	2020	10 722	3896,88
Веб- та мультимедійні розробники	2021	14 747	4062,21
Веб- та мультимедійні розробники	2022	8927	5015,54
Веб- та мультимедійні розробники	2023*	1590	7434,84

Джерело: розрахунки автора за даними бази *ETF*; *дані в процесі скрейпінгу та структурування, тому інформація є неповною.

Використання *Snowflake* в контексті аналізу великих даних *ETF* є прикладом ефективного застосування сучасних технологічних рішень для глибокого розуміння та адаптації до динамічних умов ринку праці. Це демонструє важливість інтеграції передових аналітичних інструментів та технологій у процеси дослідження та розробки політик, спрямованих на підвищення ефективності ринку праці та реагування на його поточні та майбутні виклики. Інформація, отримана в ході аналізу даних *ETF* за допомогою *Snowflake*, може бути використана для підтримки рішень у сфері політики зайнятості.

Спираючись на результати аналізу бази даних *ETF*, можна помітити, що деякі професії на сучасному ринку праці України стикаються зі зменшенням попиту на фахівців у сфері бізнес-адміністрування та комерції, і це є важливим індикатором ширших соціально-економічних змін (табл. 2). Така тенденція може бути пов'язана з наслідками воєнної агресії з боку РФ, яка викликала серйозні зміни у різних сферах професійної діяльності, включаючи широкомасштабне переміщення бізнесів. Водночас сфера інформаційних технологій (ІТ) продовжує утримувати стабільну кількість вакансій, порівняно з попередніми роками, відзначається зростання середнього рівня заробітних плат у цій галузі. Це вказує на зростаючий попит на висококваліфікованих фахівців у сфері ІТ, що може бути зумовлено збільшенням потреби в інформаційних ресурсах та технологіях. Тенденція відображає зміни на ринку, де якість та спеціалізація працівників стають вирішальними у контексті цифрової трансформації.

Швидке зростання сектору ІТ в Україні, підтверджене результатами аналізу даних з онлайн-порталів пошуку роботи, вказує на збільшення попиту на висококваліфікованих фахівців у цій галузі. Це зумовлено не лише

Таблиця 2. Вибіркове порівняння професій в сфері бізнес-адміністрування та комерції за останні 4 роки

	Вакансії за роками				Усього вакансій за 4 роки
	2020	2021	2022	2023*	
Бухгалтери	2455	2123	805	360	5743
Молодші спеціалісти з бухгалтерського обліку	11 571	11 693	8418	10 380	42 062
Адміністративні та виконавчі секретарі	2745	2688	811	974	7218
Фахівці з реклами та маркетингу	2255	2344	1616	1249	7464
Менеджери з реклами та зв'язків з громадськістю	2117	2052	551	200	4920
Комерційні торгові представники	5312	4798	2060	3672	15 842

Джерело: розрахунки автора за даними бази *ETF*; * дані в процесі скрейпінгу та структурування, тому інформація є неповною

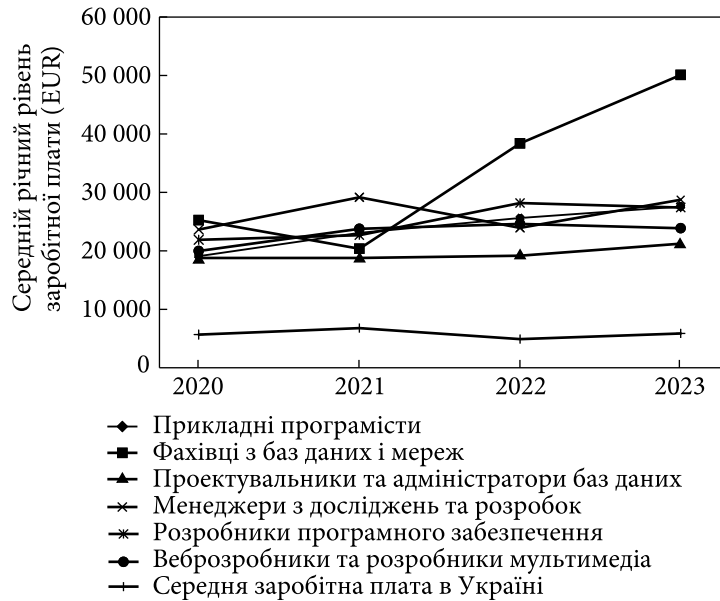


Рис. 3. Оголошена динаміка заробітних плат в ІТ-секторі за даними сайтів вакансій
Джерело: розрахунки автора за даними бази ЕТФ. Для порівняння наведена річна середня заробітна плата в Україні за обраними роками за середньобанківським валютним курсом.

потребами внутрішнього ринку, але й глобальними трендами, які диктують активну цифровізацію та автоматизацію бізнес-процесів. Важливим фактором, який підсилює цю тенденцію, є стрімкий розвиток технологічних інновацій, зокрема в таких областях, як штучний інтелект, машинне навчання, розробка програмного забезпечення та кібербезпека. Крім того, зростання чисельності працівників у інноваційних галузях (наприклад, в ІТ), свідчить про активну адаптацію України до викликів глобальної цифрової економіки. Це зростання відображає зусилля країни у відповіді на потребу в нових навичках та компетенціях, необхідних для розвитку сучасного інформаційного сектору.

У сфері інформаційних технологій в Україні, разом зі зростанням попиту на кваліфікованих фахівців, спостерігається також значне покращення фінансових показників, зокрема в розмірі заробітних плат (рис. 3). Ця позитивна динаміка є результатом зростаючої вартості ІТ-навичок у глобальній цифровій економіці та відображає збільшення внеску галузі в економіку країни. Високі зарплати в ІТ-секторі є віддзеркаленням не тільки підвищеного попиту на ці спеціалізації, але й визнання значення цих професій у сучасному технологічному світі.

Заробітні плати в ІТ-галузі часто значно перевищують середні показники по країні, що зумовлено високою конкуренцією за кваліфіковані

кадри, а також необхідністю винагороджувати складні та високотехнологічні навички, які є критичними для успіху в галузі. Також це свідчить про інвестиції бізнесу в розвиток своїх технологічних потужностей, адже кваліфіковані ІТ-фахівці відіграють ключову роль у реалізації інноваційних проєктів та цифровій трансформації бізнесу.

Такі зміни на ринку праці сприяють розвитку стартап-екосистеми в Україні, де інноваційні ідеї та проєкти мають можливість отримати підтримку та інвестиції. Це стимулює економічний розвиток країни, сприяє створенню нових робочих місць та посилює її конкурентоспроможність на міжнародній арені.

Отже, динаміка ринку праці України в області ІТ та інших інноваційних секторів є яскравим свідченням того, як країна адаптується до змінюваних умов глобального ринку, розвиваючи технологічні навички та сприяючи економічному прогресу.

Висновки. Для сучасного ринку праці можливість використання та значення релевантних великих даних невинно зростає, стаючи фундаментальним джерелом для розробки інформаційних систем та стратегій. База даних, сформована *ETF*, уміщує великий обсяг даних про вакансії в Україні за останні п'ять років і постає важливим джерелом інформації для аналізу та прогнозування майбутніх тенденцій, відображаючи динаміку попиту на актуальні професії та навички.

Використання сучасних технологій у сфері збору великих даних (веб-скрейпінг та парсинг), відіграє критичну роль у підвищенні ефективності досліджень ринку праці. Ці методи забезпечують можливість ефективного формування великих обсягів даних щодо поточного стану ринку праці.

Платформа *Snowflake*, разом з іншими подібними сервісами, надає інструменти для обробки та аналізу великих даних. Завдяки унікальній архітектурі, яка дає змогу незалежно масштабувати обчислювальні та зберігаючі ресурси, *Snowflake* забезпечує високу продуктивність і оптимальне управління даними. Це значно полегшує процес аналізу великих обсягів інформації, дозволяючи дослідникам та аналітикам отримувати важливі інсайти з даних, що є неоціненним для розуміння та прогнозування тенденцій ринку праці.

Подальші дослідження повинні зосередитись на розширенні можливостей аналізу великих даних, зокрема через розробку нових моделей машинного навчання та штучного інтелекту, що можуть допомогти в адаптації стратегій зайнятості до швидкозмінних умов ринку. Такі дослідження мають на меті не лише вдосконалення інструментів аналізу, але й розробку комплексних рекомендацій для політики зайнятості, що відповідає потребам сучасного ринку праці.

Зростання попиту на фахівців у сфері інформаційних технологій та високі показники заробітної плати в ІТ-секторі в Україні вказують на збіль-

шення цінності ІТ-навичок у глобальному масштабі. Це підкреслює необхідність інвестицій у розвиток технологічних потужностей та підготовку кадрів, які відповідають сучасним вимогам ринку праці.

Слід констатувати, що інтеграція сучасних інструментів та технологій збору та обробки даних щодо стану ринку праці необхідна для підвищення ефективності політики регулювання ринку праці та ринку освітніх послуг, а також їх адаптації до сучасних динамічних умов. Важливим аспектом подальших досліджень є розуміння соціально-економічного впливу великих даних на ринок праці, аналіз змін у навичках та кваліфікаціях, необхідних для новітніх професій, та розробка стратегій для підвищення гнучкості та інклюзивності ринку праці.

ЛІТЕРАТУРА

1. Mashey J.R. Big data and the next wave of {InfraStress} problems, solutions, opportunities. In: *1999 USENIX annual technical conference (USENIX ATC 99)*. 1999. URL: https://static.usenix.org/event/usenix99/invited_talks/mashey.pdf (дата звернення: 24.01.2024).
2. Sarioglu V., Ogay M. Approach to population estimation in Ukraine using mobile operators' data. *Statistics in Transition new series*, 2023. Т. 24 (1). P. 131—144. <https://doi.org/10.59170/stattrans-2023-008>
3. Chetty R., Hendren N. The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates. *Harvard University and NBER*, 2015. Vol. 133 (3). P. 14—16.
4. Sarioglu V., Cymbal O. Labour market landscaping Ukraine. *Big Data for labour market intelligence*. 2020. 29 p.
5. Vaccarino A. Big Data For Labour Market Information (Lmi) In Ukraine. Methodological overview and analytics insights on Ukrainian Web Labour Market Working Paper. 2020. P. 6—33. URL: https://www.etf.europa.eu/sites/default/files/2021-05/ukraine_big_data_lmi_analysis_2020_web.pdf (дата звернення: 24.01.2024).
6. Vaccarino A. Big Data For Labour Market Information (Lmi) In Tunisia. Methodological overview and Analytics insights on Tunisian Web Labour Market. Working Paper. 2020. P. 11—30. URL: https://www.etf.europa.eu/sites/default/files/2021-05/tunisia_big_data_lmi_analysis_2020_web.pdf (дата звернення: 24.01.2024).
7. Boughzala M. Big Data for labour market intelligence. Web labour market of Tunisia. Landscaping and brief overview. 2020. P. 3—6. https://www.etf.europa.eu/sites/default/files/2020-11/big_data_lmi_tunisia_landscaping_ojv_sources_2020_web.pdf (дата звернення: 24.01.2024).
8. Dusi S., Graziani C., Mezzanzanica M., Pelucchi M. Informal Employment: Can Big Data Improve Knowledge on This Phenomenon? *Assessing Informal Employment and Skills Needs*. Rainer Hampp Verlag, 2019. P. 197—230. <https://doi.org/10.5771/9783957103598-197>
9. Giambona F., Khalawi A., Buzzigoli L., Grassini L., Martelli C. Big data analysis and labour market: an analysis of Italian online job vacancies data. In: *ASA 2021 Statistics and Information Systems for Policy Evaluation: book of short papers of the on-site conference*. Firenze University Press, 2021. P. 117—120.
10. Офіційна бібліотека XML Path Language (XPath). URL: <https://www.w3.org/TR/xpath-10/> (дата звернення: 24.01.2024).
11. Офіційна бібліотека мови Python. URL: <https://docs.python.org/3/faq/library.html> (дата звернення: 24.01.2024).

12. Thapelo T.S., Namoshe M., Matsebe O., Motshegwa T., Bopape M.-J.M. SASSCAL WebSAPI: A Web Scraping Application Programming Interface to Support Access to SASSCAL's Weather Data. *Data Science Journal*. 2021. Vol. 20. P. 5. <https://doi.org/10.5334/dsj-2021-024>
13. Chamberlin D. D., Boyce R. F. SEQUEL: A Structured English Query Language. Proceedings of the 1974 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control (SIGFIDET' 74). 1974. ACM Press, Ann Harbor. P. 249—264. URL: <https://s3.us.cloud-object-storage.appdomain.cloud/res-files/2705-sequel-1974.pdf> (дата звернення: 24.01.2024).

REFERENCES

1. Mashey, J. R. (1999). Big data and the next wave of {InfraStress} problems, solutions, opportunities. *1999 USENIX annual technical conference (USENIX ATC 99)*. https://static.usenix.org/event/usenix99/invited_talks/mashey.pdf
2. Sarioglo, V., & Ogay, M. (2023). Approach to population estimation in Ukraine using mobile operators' data. *Statistics in Transition new series*, 24 (1), 131—144. <https://doi.org/10.59170/stattrans-2023-008>
3. Chetty, R., & Hendren, N. (2015). The impacts of neighborhoods on intergenerational mobility: Childhood exposure effects and county-level estimates. *Harvard University and NBER*, 133 (3), 14—16.
4. Sarioglo, V., & Cymbal, O. (2020). Labour market landscaping Ukraine. *Big Data for labour market intelligence*, 29 p.
5. Vaccarino, A. (2020). Big Data For Labour Market Information (Lmi) In Ukraine. Methodological overview and analytics insights on Ukrainian Web Labour Market Working Paper, 6—33. https://www.etf.europa.eu/sites/default/files/2021-05/ukraine_big_data_lmi_analysis_2020_web.pdf
6. Vaccarino, A. (2020). Big Data For Labour Market Information (Lmi) In Tunisia. Methodological overview and Analytics insights on Tunisian Web Labour Market. Working Paper, 11—30. https://www.etf.europa.eu/sites/default/files/2021-05/tunisia_big_data_lmi_analysis_2020_web.pdf
7. Boughzala, M. (2020). Big Data for labour market intelligence. Web labour market of Tunisia. Landscaping and brief overview, 3—6. https://www.etf.europa.eu/sites/default/files/2020-11/big_data_lmi_tunisia_landscaping_ojv_sources_2020_web.pdf
8. Dusi, S., Graziani, C., Mezzananza, M., & Pelucchi, M. (2019). Informal Employment: Can Big Data Improve Knowledge on This Phenomenon? *Assessing Informal Employment and Skills Needs*, 197—230. Rainer Hampp Verlag. <https://doi.org/10.5771/9783957103598-197>
9. Giambona, F., Khalawi, A., Buzzigoli, L., Grassini, L., & Martelli, C. (2021). Big data analysis and labour market: an analysis of Italian online job vacancies data. *ASA 2021 Statistics and Information Systems for Policy Evaluation: book of short papers of the on-site conference, Vol. 132*, 117—120. Firenze University Press.
10. Official XML Path Language (XPath) Library (2023). <https://www.w3.org/TR/xpath-10/>
11. Official library of the Python language (2023). <https://docs.python.org/3/faq/library.html>
12. Thapelo, T. S., Namoshe, M., Matsebe, O., Motshegwa, T., & Bopape, M.-J. M. (2021). SASSCAL WebSAPI: A Web Scraping Application Programming Interface to Support Access to SASSCAL's Weather Data. *Data Science Journal*, 20, 5. <https://doi.org/10.5334/dsj-2021-024>

13. Chamberlin, D. D., & Boyce, R. F. (1974). SEQUEL: A Structured English Query Language. Proceedings of the 1974 ACM SIGFIDET (Now SIGMOD) Workshop on Data Description, Access and Control (SIGFIDET' 74), ACM Press, Ann Harbor, 249—264. <https://s3.us.cloud-object-storage.appdomain.cloud/res-files/2705-sequel-1974.pdf>

Стаття надійшла до редакції журналу 09.01.2024

Mykhailo Rozbytskyi, Chief economist
Institute for Demography and Life Quality Problems of the NAS of Ukraine
01032, Ukraine, Kyiv, Blvd Tarasa Shevchenka, 60
E-mail: rozbytskyi.m@gmail.com
ORCID: 0009-0007-4102-0912

USAGE OF BIG DATA FOR INFORMATION SUPPORT OF THE LABOR MARKET

The challenges of collecting current labor market data amidst global changes and geopolitical fluctuations increasingly necessitate new approaches and alternative methods of data collection and analysis. This highlights the relevance of research aimed at developing and adapting labor market information provision methods to contemporary challenges. A promising approach in this context is the use of Big Data for labor market assessment, which involves collecting information from sources such as online job search and vacancy portals. This method allows for a deeper analysis of market trends and provides a more accurate and timely assessment of labor market needs and opportunities. The aim of this article is to discuss approaches to developing labor market information systems using Big Data, particularly online data from job vacancy websites. It examines the use of Big Data in labor market analysis based on a database containing over four million job vacancies posted on Ukrainian job search portals over the last five years, provided by the European Training Foundation (ETF). The effectiveness of these approaches in facilitating job search for all interested parties is evaluated, particularly through providing insights into the dynamics of supply and demand in the labor market based on data from these portals. The opportunities and limitations of using Big Data in this context are analyzed, including their impact on employment policy development and labor market planning. The potential benefits of Big Data in providing deeper and more accurate market condition analyses are outlined, along with technical aspects and challenges associated with their processing and interpretation. The article examines methodological approaches to data collection and analytical processing in the context of accelerated transformations, volatility, and limited access to traditional information resources. The scientific novelty of the article lies in the substantiation of the feasibility and appropriateness of using open data from online job portals for labor market information provision under current conditions. In conducting the research, methods of analysis, synthesis, and generalization were applied to identify the main contemporary issues of labor market information provision in Ukraine. The effectiveness of data collection methods based on web scraping and parsing techniques was evaluated, as well as the use of the integrated Snowflake platform to identify key trends and patterns in the labor market. The conclusions summarize the main points and substantiate directions for further research, highlighting the significance of Big Data in developing employment strategies and optimizing the labor market.

Keywords: big data, labor market, web scraping, parsing, employment, trends, vacancies, demand.