

Гладка Ю.А., к.фіз.-мат.н., доцент,
доцент кафедри комп'ютерної математики та інформаційної безпеки,
Київський національний економічний університет
імені Вадима Гетьмана

Дубецький О.В., магістр спеціальності «Системний аналіз»,
кафедра комп'ютерної математики та інформаційної безпеки,
Київський національний економічний університет
імені Вадима Гетьмана

Gladka Yu.A., candidate of Physical and Mathematical Sciences, Associate
Professor, Department of Computer Mathematics and Information Security,
KNEU named after Vadym Hetman

Dubetsky O.V., master's degree in «System Analysis»,
Department of Computer Mathematics and Information Security,
KNEU named after Vadym Hetman

СИСТЕМНИЙ АНАЛІЗ ВПЛИВУ ЦИФРОВИХ КОМПЕТЕНЦІЙ НА РИНОК ПРАЦІ УКРАЇНИ

SYSTEM ANALYSIS OF THE IMPACT OF DIGITAL COMPETENCES ON THE LABOR MARKET OF UKRAINE

Анотація. Інтелектуальний аналіз даних є однією з ефективних сфер створення інноваційних даних, призначених для виявлення прихованої та цінної інформації. Досягнення наукової інформаційної експертизи застосовуються у різних сферах людської діяльності, однією з таких сфер є ринок праці. У статті досліджено питання впливу цифрових компетенцій на ринок праці України. Проведено інтелектуальний аналіз даних ринку праці на прикладі вакансій. Встановлено поточний стан використання цифрових компетенцій в Україні. Виявлено проблемні аспекти аналітичної інтерпретації цифрових компетенцій та забезпечення повноти розкриття інформації у сфері цифрових компетенцій в умовах цифрових трансформацій, запропоновано шляхи вирішення. Доведено, що зіставлення тегів на основі правил відповідає всім необхідним навичкам, але також виділяє інші іменники та власні іменники. Через великий шум і для вивчення використано інший словник для отримання необхідних іменників (навичок). Цю проблему вирішено за допомогою інших моделей NLP, таких як тематичне моделювання, LSTM і будовування слів. Зроблено висновок з аналізу, що вилучення контекстних тем без шуму усе ще вивчається. Проведене дослідження дозволило визначити низький рівень застосування цифрових компетенцій у посадових вимогах. Підтверджено, що це свідчить про недостатній розвиток цифровізації і пов'язано безпосередньо з економічною ситуацією в країні, яка має місце обмеженість ресурсів.

Ключові слова: системний аналіз, інтелектуальний аналіз даних, цифрова компетентність, цифрові технології, вирішення проблем, цифровізація, цифрові перетворення, ефективність.

Abstract. Data mining is one of the effective areas of creating innovative data designed to uncover hidden and valuable information from data. Achievements of scientific information expertise find their application in various spheres of human activity, one of such spheres is the labor market. The article examines the impact of digital competencies on the labor market of Ukraine. An intellectual analysis of labor market data was carried out using the example of vacancies. The current state of use of digital competencies in Ukraine is established. Problematic aspects of the analytical interpretation of digital competences and ensuring the completeness of information disclosure in the field of digital competences in the conditions of digital transformations are identified, solutions are proposed. Rule-based tagging meets all the necessary skills, but also highlights other nouns and proper nouns. There was a lot of noise and mining required another dictionary to get the required nouns (skills). This problem is interesting to solve using other NLP models, such as topic modeling, LSTM, and word embedding. The most important conclusion from the analysis is that noise-free context topic extraction is still under investigation. The conducted research made it possible to determine the low level of application of digital competencies in job requirements. This indicates the insufficient development of digitalization, which is directly related to the economic situation in the country, and there is a limitation of resources.

Keywords: system analysis, intelligent data analysis, digital competence, digital technologies, problem solving, digitalization, digital transformations, efficiency.

Актуальність теми дослідження. Цифрові навички сьогодні вже не є спеціальними знаннями, якими володіють лише спеціалісти з ІКТ. Це основна компетентність сучасної людини, обов'язкові навички, без яких з кожним роком стає все важче соціалізуватися, адаптуватися, навчатися чи просуватися по кар'єрних сходах.

Дедалі частіше з'являються дослідження, в яких вчені стверджують, що в найближчому майбутньому майже кожна професія буде у різний спосіб пов'язана з цифровими навичками через швидкі зміни в технологіях.

Отже, системний аналіз впливу цифрових компетенцій на ринок праці України вбачається актуальним науковим завданням.

Постановка проблеми. Інтелектуальний аналіз даних є однією з ефективних галузей створення інноваційних даних, призначених для виявлення прихованої та цінної інформації з даних. Досягнення наукової інформаційної експертизи знаходять своє застосування в різних сферах людської діяльності, однією з таких сфер є ринок праці.

Україна, як і більшість країн світової економіки, стикається з різними суттєвими змінами: прогресивні технології, старіння населення та інші, що впливає на попит і пропозицію компетенцій. Труднощі полягають у розвитку ефективності компетенцій та їх невідповідності, недостатні інвестиції та недолік ключових компетенцій. Проблемні ситуації, такі як коронавірус і військові за-

грози, впливають на потреби в здібностях, а також на пропозиції роботи. Навички змішуються та закріплюються у нові комбінації. Вимоги до компетентності змінюються настільки швидко, що більшість наборів обов'язків не відповідають сьогоденню, не кажучи вже про шукачів роботи, постачальників освітніх послуг та роботодавців.

Як зазначено Всесвітнім економічним форумом (WEF, 2019), стає більш очевидним, що ринок праці повинен створити навички як спільну валюту, підтримувати співпрацю між роботодавцями та освітянами. Слід розглянути загальний метод оновлення і консолідації таксономії компетенцій, групування навичок та їх визначення. Такий підхід має потенціал для створення фундаменту для більш ефективного ринку для підвищення кваліфікації та перекваліфікації [2].

Ідея цього аналізу полягає в тому, щоб зібрати дані з Державного центру зайнятості (<https://www.dcz.gov.ua/>) зі списками вакансій, щоб витягти навички / очікування, які часто потрібні для виконання різноманітних робіт.

Аналіз останніх досліджень і публікацій. Питання інформаційно-комунікаційної або цифрової компетентності, що дозволяють опрацьовувати інформацію, забезпечивши перехід від кваліфікаційної моделі до повноцінних компетентнісних моделей, які супроводжують людей упродовж їх особистісного розвитку і професійної кар'єри, знаходиться в полі зору як теоретиків, так і практиків. Цій проблемі приділяють значну увагу як українські (Л. Боярчук [3], М. Махсма [4]), так і зарубіжні (Л. Бревер [5], Дж. Джеймс [6]) дослідники. Але незважаючи на цінність проведених досліджень, питання цифрових компетенцій в Україні залишаються невирішеними та потребують подальшого дослідження.

Постановка завдання. Мета статті полягає в інтелектуальному аналізі сучасного стану використання цифрових навичок на ринку праці України.

Виклад основного матеріалу. Згідно з Рамками цифрової компетентності для громадян України цифрова компетентність є ключовою в умовах Четвертої промислової революції. Цей термін містить впевнене, критичне, творче та відповідальне використання і взаємодію з засобами цифрових технологій для навчання, працевлаштування, роботи, дозвілля та участі у суспільному житті. Він охоплює такі поняття, як інформаційна грамотність та медіаграмотність, комунікація та співпраця, створення цифрового контенту (включаючи основи програмування), безпека (а та-

кож захист персональних даних у цифровому середовищі та кібербезпеку), а також розв’язання різнопланових проблем з використанням цифрових технологій навчання впродовж життя. Детальний опис та рівні володіння наведено в Рамці цифрової компетентності для громадян України (Опис рамки цифрової компетентності для громадян України, 2021) [7].

Після успішного збирання та форматування даних із Державного центру зайнятості (<https://www.dcz.gov.ua/>) дані містили чіткий ідентифікатор, назву вакансії, регіон, опис роботи, дату оголошення, рівень оплати праці та назву компанії. Дані про вакансії включають період з 2020-01-04 по 2022-12-09 та налічує 800 353 вакансії з 4920 спеціалізацій (табл. 1).

Таблиця 1

ФРАГМЕНТ ФРЕЙМУ ОТРИМАНИХ ДАНИХ

	id	name	region	description	pubdate	salary	company
60380	1044882174	охоронець	Житомирська область, Коростень	заробітна плата усього грн опис ...	2020-06-17	5000	652, КОРОСТЕНЬСЬКИЙ МЦЗ
194062	1058885347	бармен	Чернівецька область, Чернівці	заробітна плата усього грн у тому чис...	2021-11-03	8000	2412, Чернівецький МЦЗ (Чернівецький регіон)
278517	1062273224	продавець продовольчих товарів	Луганська область, Біловодський район, Біловодськ	заробітна плата усього грн у тому чис...	2021-06-05	6000	1215, Біловодський районний центр зайнятості
45393	1039715874	бібліотекар	Чернівецька область, Чернівці	заробітна плата усього грн у тому чис...	2020-02-19	5500	2412, Чернівецький МЦЗ (Чернівецький регіон)
299075	1051226405	оператор копальні	Миколаївська область, Доманівський район, Цвіт...	заробітна плата усього грн у тому чис...	2020-12-10	5000	1463, ДОМАНІВСЬКА РАЙОННА ФІЛІЯ МИКОЛАЇВСЬКОГО...

Після того як були зібрані дані, ми відформатували більшість стовпців у керований спосіб. Більшість роботи, яку потрібно буде виконати, — це вилучення характеристик із опису вакансій. Як початковий крок виконано всі звичайні етапи попередньої обробки, необхідні для обробки природної мови. Це і використання регулярних виразів для видалення непотрібних символів, і видалення типових та нерелевантних стоп-слів, і маркування кожного окремого опису. Розбір тексту на слова, а також приведення їх до початкової форми виконано за допомогою бібліотеки Rymorphy2 [8].

Процес, який тут застосовувався, полягав у розбитті на фрагменти — це спосіб вилучення певних фраз із документа. Спочатку планувалось використовувати спеціальний NER (Name Entity Recognition), щоб отримати навички з опису посади. Дослідивши способи впровадження цього рішення, ми вирішали використовувати класифікацію на основі фрагментації, а не NER через кількість необхідних навчальних даних. Створення навчального та тестового наборів для NER зайняло б набагато більше часу.

Серед тих слів, які потрібно розібрати, давайте подивимося на найчастіші слова, які використовуються в описах вакансій (рис. 1).

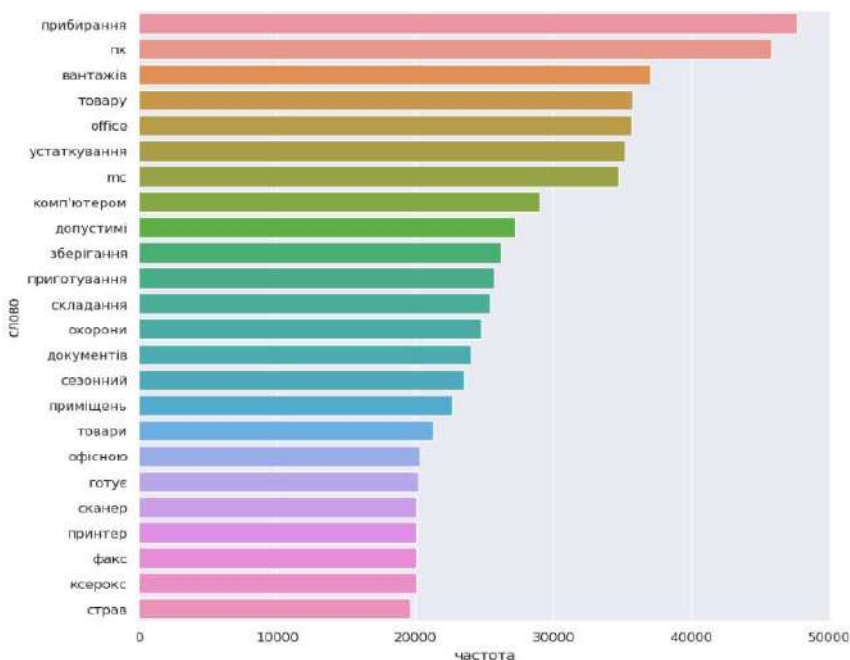


Рис. 1. Візуалізація найчастіших слів

Джерело: розроблено авторами.

Зрозуміло, що ці головні слова навряд чи є навичками. Можна очікувати, що слово «пк» буде часто зустрічатися, але навіть це слово з'являється в корпусі лише 45787 раз. Можливо, біграма (рис. 2) чи триграма (рис. 3) почнуть розкривати навички в описі вакансії? Що ж, ми бачимо деякі кращі результати, і деякі навички починають з'являтися в найчастіших даних n -грамів, але має бути більш ефективний спосіб ідентифікації цих навичок в описі посади.

Ми використати нейронні мережі для нашого другого підходу та спробували модель word2vec [9], щоб отримати необхідні вектори навичок. Ми навчили модель за допомогою набору токенів із вилученого опису вакансії та запустили вимірювання подібності. Наведений нижче фрагмент коду показує найбільш схожі слова у домені цифрових навичок (рис. 4). Word2vec також корисний для конкретних цифрових навичок.

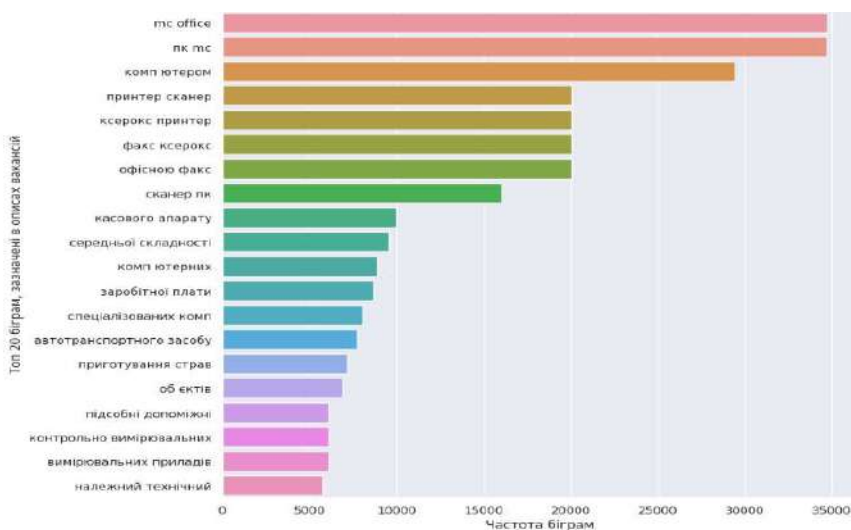


Рис. 2. Візуалізація найчастіших біграм

Джерело: розроблено авторами.

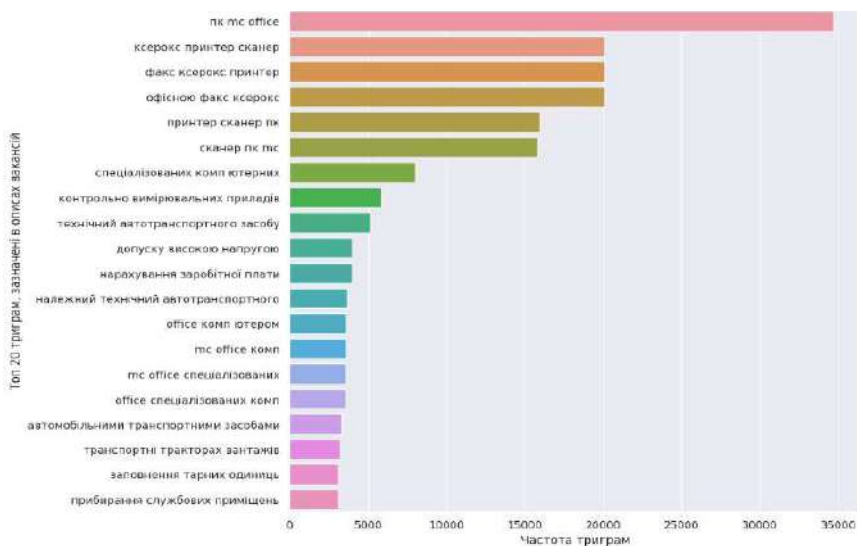


Рис. 3. Візуалізація найчастіших триграм

Джерело: розроблено авторами.

```
# digital specific skills
model.wv.most_similar([['office', 'word', 'excel', 'сканер', 'принтер', 'програмування', 'розробка',
                        'програма', 'комунікація', 'співпраця'], topn=25])
```

```
[('sql', 0.6188371777534485),
 ('excel', 0.6108712418556213),
 ('server', 0.5948401689529419),
 ('windows', 0.5821477174758911),
 ('adobe', 0.581726610660553),
 ('photoshop', 0.5815576314926147),
 ('html', 0.5787284970283588),
 ('css', 0.5703566074371330),
 ('corel', 0.5687632994651794),
 ('програма', 0.5599780882762637),
 ('draw', 0.555258572101593),
 ('erp', 0.5471287368728080),
 ('point', 0.5455071926116943),
 ('linux', 0.5435844659805298),
 ('web', 0.5325785875320435),
 ('illustrator', 0.5323614478111267),
 ('mysql', 0.5307382345199585),
 ('програми', 0.5307325720787048),
 ('autocad', 0.5301836729049683),
 ('excel', 0.5275077223777771),
 ('medoc', 0.5256964564323425),
 ('адміністрування', 0.5248759984970093),
 ('клієнт-банк', 0.519995391368866),
 ('internet', 0.5174419283866882),
 ('google', 0.5171661972999573)]
```

Рис. 4. Візуалізація найбільш схожих слів у домені цифрових навичок

Джерело: розроблено авторами.

Грунтуючись на нашому аналізі, ми визначили різні набори навичок, яким надають топ 25 назв вакансій із 1544, в яких використовуються цифрові компетенції (табл. 2). З усього проаналізованого корпусу кількість назв вакансій становить 31,38 % із цифровими компетенціями.

Таблиця 2

ТОП-25 НАЗВ ВАКАНСІЙ З ЦИФРОВИМИ КОМПЕТЕНЦІЯМИ

№ з/п	Назва вакансії	Частота у корпусі
1	бухгалтер	6257
2	спеціаліст державної служби (місцевого самоврядування)	2609
3	головний бухгалтер	1513
4	фахівець	1459
5	менеджер (управитель) із збуту	1268
6	економіст	1103
7	комірник	764
8	продавець-консультант	708
9	адміністратор	693
10	інженер	687

№ з/п	Назва вакансії	Частота у корпусі
11	вчитель закладу загальної середньої освіти	686
12	менеджер (управитель)	611
13	начальник відділу	599
14	оператор комп'ютерного набору	556
15	діловод	547
16	юрисконсульт	509
17	інженер-конструктор	496
18	секретар	487
19	інженер з охорони праці	471
20	оператор поштового зв'язку	463
21	продавець непродовольчих товарів	462
22	інспектор з кадрів	412
23	інженер-програміст	407
24	фахівець з публічних закупівель	398
25	начальник відділу поштового зв'язку	385

Висновок. У нашому інтелектуальному аналізі даних використовувалися подвійні алгоритми. В обох є застереження. Подібність word2vec базується на описі, який ми навчили, тому пошук навичок для кожної ролі не вдається, якщо в описі не згадується роль. Словниковий запас залежить від корпусу, тому рольовий пошук не був успішним із моделлю word2vec.

Зіставлення тегів на основі правил відповідає всім необхідним навичкам, але також виділяє інші іменники та власні іменники. Було багато шуму, і для видобування потрібен був інший словник для отримання необхідних іменників (навичок). Цю проблему цікаво вирішити за допомогою інших моделей NLP, таких як тематичне моделювання, LSTM і вбудовування слів. Найважливіший висновок із цього аналізу полягає в тому, що вилучення контекстних тем, без шуму, досі досліджується.

Проведене дослідження дозволило визначити низький рівень застосування цифрових компетенцій у посадових вимогах. Це свідчить про недостатній розвиток цифровізації, що безпосередньо пов'язано з економічною ситуацією в країні, має місце обмеженість ресурсів.

Подальші наукові дослідження доцільні в напрямках: розроблення методик з питань застосування цифрових компетентностей; підвищення ефективності роботи служби зайнятості через її цифрову трансформацію.

Бібліографічні посилання

1. Toby Segaran. Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly Media, Inc, 2007. 308 с.
2. World Economic Forum. (2019). Strategies for the new economy: Skills as the currency of the labour market. URL: <https://www.weforum.org/whitepapers/strategies-for-the-new-economy-skills-as-the-currency-of-the-labour-market>
3. Боярчук Л.В. Застосування зарубіжного досвіду в роботі Державної служби зайнятості України. *Науковий вісник Полісся*. № 20151(1). С. 65—70.
4. Махсма М. Світові тенденції трансформації зайнятості населення в умовах глобалізації економіки. *Україна: аспекти праці*. 2007. № 4. С. 10—15.
5. Brewer L. Enhancing youth employability: What? Why? and How? Guide to core work skills. URL: https://www.ilo.org/wcmsp5/-groups/public/@ed_emp/@ifp_skills-/documents/publication/wcms_213452.pdf
6. James J. Heckman Hard Evidence on Soft Skills. *National Bureau of Economic Research*. 2012. June. URL: <http://www.nber.org/papers/w18121>
7. Опис Рамки цифрових компетентностей для громадян України. Міністерство цифрової трансформації України. 2021. URL: https://thedigital.gov.ua/storage/uploads/files/news_post/2021/3/mintsifra-oprilyudnyue-ramku-tsifrovoi-kompetentnosti-dlya-gromadyan/%D0%9E%D0%A0%20%D0%A6%D0%9A.pdf
8. PyMorphy2. URL: <https://pymorphy2.readthedocs.io/en/stable/>
9. URL: <https://code.google.com/archive/p/word2vec/>

Статтю подано до редакції 26.11.2022