

В спеціалізовану вчену раду К 29.051.16
при Східноукраїнському національному університеті ім. В. Даля,
м. Сєвєродонецьк, пр. Центральний, 59-а

ВІДГУК

офіційного опонента доктора технічних наук, доцента

Доценка Сергія Ілліча

на дисертаційну роботу

Барбарук Ліни Вікторівни на тему:

«Моделі та метод обробки великих даних в інформаційно-аналітичних системах моніторингу водних об'єктів»,
подану на здобуття наукового ступеня кандидата технічних наук
за спеціальністю 05.13.06 – інформаційні технології

Актуальність теми. Найважливішим завданням людства є збереження довкілля – основи безпечної існування та здоров'я планети. Вода є наріжним каменем життя, а її якість є одним з найважливіших показників стану навколишнього середовища. Це життєво важлива основа майже кожного сектора світової економіки, будь то сільськогосподарський, промисловий, комерційний або рекреаційний. Тому моніторинг водних ресурсів, як системи безперервного (поточного) і комплексного відстеження стану водних об'єктів, контролю й обліку їх кількісних і якісних характеристик в часі, є вирішальним завданням у всьому світі. Своєчасне виявлення і прогнозування розвитку негативних процесів, що впливають на якість води у водних об'єктах і їх екологічний стан, розробка і реалізація заходів щодо запобігання негативним наслідкам цих процесів і є головним завданням моніторингу водних ресурсів навколишнього природного середовища. Моніторинг якості води в навколишньому середовищі має на меті надати дані, необхідні для захисту навколишнього середовища від несприятливих біологічних наслідків багаторазового хімічного забруднення, спричиненого антропогенними дифузними викидами та точковими джерелами.

Водночас стан якості води є результатом складних природних і техногенних умов та наслідків взаємодії як у часі, так і в просторі. Отже, абстрагувати суть умов якості води часто буває дуже складно. Крім того, відповідно до стандартів якості поверхневої води, при моніторингу, аналізі та прийнятті рішень щодо якості води оцінка якості води може вимагати різноманітного набору технологій та датчиків. Очевидно, що аналіз та обробка багатовимірних даних із мережі моніторингу є типовою проблемою обробки великих даних. У дисертаційній роботі Барбарук Ліни Вікторівни подано вирішення цього актуального наукового завдання шляхом створення науково-методичних основ і математичного апарату інформаційних систем

здатних забезпечити ефективну обробку та використання великих даних, отримуваних від систем моніторингу водних об'єктів.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження проводилося на кафедрі комп'ютерних наук та інженерії Східноукраїнського національного університету ім. Володимира Даля у рамках наступних науково-дослідних робіт: “Дослідження стратегій та механізмів прийняття рішень для інтегрованого управління водними ресурсами” (Східноукраїнський національний університет імені Володимира Даля, № ДР 0116U005784, 12.2016-07.2020 pp.); “Проектування системи моніторингу та контролю водних об'єктів на основі технології інтернет речей”, (№ ДР 0120U100421, 01.2020-12.2023 pp.); Проект Європейського Союзу ERASMUS+ ALIOT 573818-EPP-1-2016-1-UK-EPPKA2-CBHE-JP “Internet of Things: Emerging Curriculum for Industry and Human Applications” (2016-2020 pp.).

Оцінка змісту дисертації, її завершеності у цілому. Дисертація складається зі вступу, чотирьох розділів, висновків і додатків.

У *вступі* зазначена необхідність проведення теоретичних та експериментальних досліджень та актуальність теми дисертаційної роботи, формулюються мета і завдання дослідження, визначається предмет, об'єкт і методи досліджень, окреслюється наукова новизна отриманих результатів, практичне впровадження та публікації.

У *першому* розділі розглянуто основні моделі та методи обробки великих даних, розглянуто потенціал інноваційних технологій з моніторингу водних об'єктів, проаналізовано особливості великих даних в системах моніторингу. Виконано огляд проблем візуалізації великих даних та підходів до їх вирішення. За результатами аналізу сформульоване основне науково-прикладне завдання роботи, яке полягає в розробці моделей та методу інформаційної технології для обробки великих даних в системах моніторингу водних об'єктів. Визначені задачі дослідження, визначено стратегію та обґрунтована методика і математичний апарат досліджень. За результатами аналізу сформульована основна стратегія процесу досліджень та визначені завдання досліджень.

Другий розділ присвячений розробці теоретичних основ та розробленню нечітких моделей для оцінювання якості води на прикладі поверхневих вод та вод рибогосподарського призначення. Виконано побудову функцій належності термів лінгвістичних змінних нечіткої моделі аналізу якості вод, розроблено моделі комплексної оцінки поверхневих вод, основою яких є відповідні функції належності.

У *третьому розділі* виконано розробку методу обробки великих даних для водойм рибогосподарського призначення. Розглянуто проблему інтерпретації результатів кластеризації на випадок застосування евристичних алгоритмів кластеризації для інтуїтивістської нечіткої обробки даних. Наведено приклад використання методу для аналізу якості вод

водоїм рибогосподарського призначення. З метою удосконалення наявних підходів до кластеризації великих даних здобувачка запропонувала узагальнення процедури автоматичного маркування нечітких кластерів, отриманих за допомогою евристичних алгоритмів кластеризації для інтуїтивістських нечітких множин.

Четвертий розділ присвячений розробці засобів та інформаційної технології для обробки великих даних, отриманих від станцій моніторингу водних об'єктів. У розділі наведені приклади практичної реалізації запропонованих у роботі моделей і методів інформаційної технології, надано архітектуру сховища даних, визначено критерії якості видобутку даних, представлено структуру бази даних та основні інструменти для аналітичної обробки даних у вигляді гібридного інструмента інтелектуального аналізу даних та управління непрямими знаннями баз даних спеціалізованої аналітичної системи водних об'єктів. Окремо виділено і сформульовано задачу візуалізації великих даних. Запропоновані рішення представлено у вигляді хордових діаграм і методів спрошення полігональних ланцюгів.

У **висновках** викладено найбільш важливі наукові й практичні результати, отримані у дисертаційній роботі, які дають розв'язок сформульованих завдань дисертаційного дослідження.

Зміст поданої дисертаційної роботи, публікацій та документів впровадження дозволяє визначити наступні положення.

Новизна наукових положень, висновків і рекомендацій. У процесі розв'язання поставлених задач автором особисто отримано ряд важливих результатів. Науковий доробок автора полягає в розвитку і поглибленні теорії та практики створення інформаційних технологій для підтримки прийняття рішень при забезпеченні процесів обробки великих даних в інформаційно-аналітичних системах моніторингу водних об'єктів і складається з наступних компонентів..

1) вперше розроблено модель для аналітичної обробки великих даних системи моніторингу водних об'єктів на основі формалізації її атриутів та інтерпретації невизначеності оцінки якості води у вигляді лінгвістичних змінних, що дозволило надати інтегровану характеристику стану водних об'єктів, для подальшого прийняття рішення;

2) удосконалено метод нечіткої кластеризації, в якому, на відміну від існуючих, виконано узагальнення процедури автоматичного маркування нечітких кластерів, отриманих за допомогою евристичних алгоритмів для інтуїтивістських нечітких даних, що дозволило застосовувати автоматичну розмітку при обробці великих даних і покращити конвергенцію алгоритму кластеризації;

3) удосконалено технологію та засоби візуалізації великих даних шляхом використання хордових діаграм та технології спрошення полігональних ланцюгів, що дозволяють зберегти баланс між вимогами до

деталізації вихідного зображення та отриманням мінімальної кількості точок на екрані систем онлайн моніторингу водних об'єктів, знижуючи час на пошук невідповідностей і ретроспективний аналіз великих даних;

4) набула подальшого розвитку інформаційна технологія обробки великих даних в інформаційно-аналітичних системах моніторингу водних об'єктів шляхом її адаптації до завдань контролю та управління рибогосподарських підприємств, що забезпечує семантичну основу для комплексної автоматизації водних господарств у частині реалізації основних аналітичних функцій.

Ступінь обґрутованості наукових положень, висновків і рекомендацій. Достовірність отриманих в дисертаційній роботі результатів забезпечується коректністю постановок задач, повнотою формулювання умов, в яких вони розв'язуються та необхідними припущеннями і обмеженнями щодо застосування результатів; використанням сучасного математичного апарату та програмного забезпечення. Авторка дисертації здійснила досить повний і всебічний аналіз розробленої проблеми, комплексне теоретичне і практичне опрацювання основних аспектів її вирішення. Наукові положення, рекомендації та висновки, сформульовані в роботі є достатньо аргументованими, містять глибокі напрацювання щодо запропонованих науково-практичних рішень. Це підтверджується перевіркою їх за критеріями вітчизняної та зарубіжної науки й практики, а також використанням ряду фундаментальних положень теорії систем, теорії прийняття рішень, теорії нечітких множин, реалізації системно-цілісного підходу до дослідження; кількісним і якісним аналізом значного обсягу теоретичного матеріалу; позитивним ефектом, отриманим в результаті практичної реалізації та впровадження розроблених моделей і методів. Коректність отриманих практичних результатів також підтверджено експериментами та чисельними розрахунками. Кожне наукове положення дисертації, що міститься у висновках, є достатньо обґрутованим в основному тексті і підтверджується апробацією результатів досліджень в практиці.

Практична значущість та можливі шляхи використання результатів дисертаційних досліджень. Основні наукові результати дисертаційної роботи мають як суто практичне, так і загальне значення, що дозволяє використовувати їх при впровадженні сучасних засобів підтримки прийняття рішень. Одержані у роботі результати є основою для обробки великих даних в системах моніторингу водних об'єктів і підтримки прийняття природоохоронних рішень. Практична цінність роботи полягає у тому, що: - розроблено елементи інформаційно аналітичної системи яка може використовуватися для накопичення екологічних даних діяльності підприємств аквакультури та водного господарства, видобутку знань з даних моніторингу водного середовища, та дозволяє зменшити час на прийняття рішень та підвищити якість рішень в системах управління

водним господарством, зокрема виробництвом рибної продукції у повністю або частково контролюваних умовах; - обґрутовано і подано варіант архітектури сховища даних, конфігурації та схеми взаємодії компонент системи підтримки прийняття рішень. Таким чином, отримані у роботі результати мають важливе науково-практичне значення у галузі побудови систем підтримки прийняття рішень в управлінні водними об'єктами. Результати роботи впроваджені в Управлінні Державного агентства рибного господарства у Луганській області. Досягнуті результати мають не тільки практичну спрямованість, але й використовуються в навчальному процесі в навчальних курсах Східноукраїнського національного університету.

Рекомендації щодо використання наукових і прикладних результатів дисертації. Підготовлені дисертантом висновки, пропозиції і рекомендації можуть бути використані на підприємствах аквакультури та водного господарства так і в організаціях, які займаються проблемами водопідготовки та водопостачання, органах виконавчої влади, вченими та викладачами вищих навчальних закладів. Результати дослідження можуть бути впроваджені на підприємствах з перевірки забрудненості води, у роботі індустріального рибництва, в системі рибного господарства та системах аквакультури тощо.

Зауваження до дисертації

1) у Розділі 1 недостатньо висвітлено питання аналізу нормативної бази щодо використання великих даних в системах моніторингу водних об'єктів, зокрема систем моніторингу на основі Інтернету речей.

2) у Розділі 2 зазначається що «...Нечіткі правила для досліджуваної водної екосистеми були розроблені за допомогою знання експертів та ретельного розгляду параметрів отриманих від системи моніторинга...», разом з тим не наведено жодної інформації щодо процедур збору та обробки даних, отриманих від експертів.

3) Результати попереднього підрахунку кількості кластерів для набору даних якості вод, візуалізовані на рис. 3.2 вказують на те, що 5 кластерів є оптимальним варіантом, проте, для подальшої роботи здобувачкою обрано 3 кластери, що потребує додаткових пояснень.

4) у Розділі 4 не достатньо висвітлено інструменти аналітичної обробки даних, зокрема взаємодія наявних методів Data Mining та розробка змішаних методів, що можуть співпрацювати з існуючими підходами для вилучення знань, що містяться в даних.

Треба зазначити, що оформлення дисертаційної роботи виконано акуратно, послідовно, логічно, але в тексті автореферату та дисертації зустрічаються описки, зокрема в авторефераті використовуються без пояснень терміни та позначення, які не є загальновідомими. Також у перелік термінів та позначень дисертації внесено не усі позначення, що використовуються у тексті дисертації, але не є загальноприйнятими. Вказані вище зауваження не мають принципового значення, скоріше носять

методичний та технічний характер, і не зменшують наукову та практичну цінність результатів роботи.

Повнота викладення наукових положень, висновків і рекомендацій в опублікованих роботах. Результати дисертації знайшли відбиття у 18 наукових публікаціях, у тому числі 8 працях у спеціалізованих фахових виданнях України, включених до міжнародних наукометрических баз, 1 стаття у закордонному виданні, що внесено до міжнародної наукометричної бази Scopus; 1 стаття у фаховому виданні іншої держави, що входить до Європейського Союзу, 8 тез в збірниках матеріалів науково-практичних конференцій. Опубліковані автором роботи повною мірою висвітлюють наукові положення, що виносяться на захист, а також отримані практичні результати. Основні наукові результати доповідалися на 12 міжнародних і всеукраїнських конференціях (2011 – 2020 рр.), обговорені і схвалені на численних наукових семінарах.

Відповідність дисертації прийнятим вимогам. Дисертація та її автореферат повністю відповідають вимогам пунктів 9, 11, 12 «Порядку присудження наукових ступенів», затверджених постановою Кабінету Міністрів України від 24 липня 2013 р. № 567 (зі змінами від 30.12.2015 та 27.07.2016 р.), що висуваються до кандидатських дисертацій.

Відповідність теми і змісту дисертації паспорту спеціальності, за якою вона подана на захист. Тема дисертації Барбарук Л.В. та її зміст відповідають формулі й галузі досліджень паспорта спеціальності 05.13.06 – інформаційні технології.

Важливість для науки і народного господарства отриманих результатів. Використання моделей та інформаційної технології обробки великих даних в інформаційно-аналітичних системах моніторингу водних об'єктів, зокрема у виробничих процесах рибогосподарських підприємств, дозволило зробити висновок, про їх ефективність в частині підвищення точності інтегральної оцінки якості вод, зменшення часу на прийняття рішень щодо якості вод водойм рибогосподарського призначення та зменшення часу на пошук невідповідностей і ретроспективний аналіз великих даних.

Висновок. На підставі вивчення дисертації, автореферату та наукових праць здобувача, опублікованих за темою дисертації, наукові результати, отримані при її виконанні, слід визнати позитивними.

Дисертаційна робота Барбарук Л.В. «Моделі та метод обробки великих даних в інформаційно-аналітичних системах моніторингу водних об'єктів» є кваліфікаційною науковою працею, що містить нові науково-прикладні результати проведених автором досліджень, що в сукупності розв'язують актуальну наукову задачу в галузі створення

інформаційних технологій, виконаним на актуальну тему, у якому отримані нові рішення зі створення інформаційних технологій для підтримки прийняття рішень при удосконаленні аналізу великих даних в системах моніторингу водних об'єктів.

Щодо мети роботи, постановки завдань і спрямованості дисертаційна робота цілком відповідає спеціальності 05.13.06 – інформаційні технології.

Основні положення дисертаційної роботи ідентичні змісту автореферату.

Стиль викладання дисертації й автореферату відповідає поставленим вимогам.

Таким чином, дисертація має наступні кваліфікаційні ознаки:

1) робота є закінченим комплексним науковим дослідженням, оформленним відповідно до вимог МОН;

2) робота містить обґрунтовані теоретичні і практичні рішення, впровадження яких дає значний внесок у розвиток інформаційних технологій і засобів підтримки прийняття рішень у питаннях аналізу великих даних в системах моніторингу водних об'єктів;

3) результати, отримані автором є новими, раніше не захищеними;

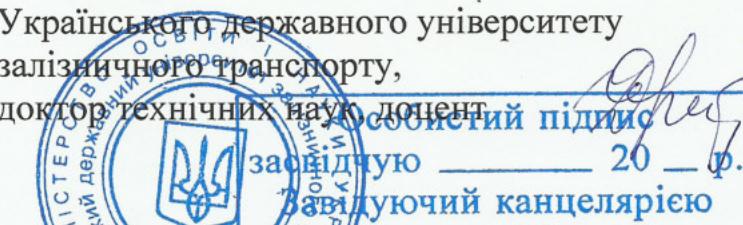
4) тема дисертації входить у тематичні плани і програми міністерств України;

5) використання результатів здійснено в діючому закладі з управління водним господарством, науковій та навчальній роботі.

Викладене вище дозволяє зробити висновок, що дисертаційна робота Барбарук Л.В., в якій вирішено актуальне науково-практичне завдання підвищення ефективності роботи інформаційно-аналітичних систем моніторингу водних об'єктів завдяки розробці та практичному застосуванню моделей та методів інформаційної технології обробки великих даних, відповідає пунктам 9, 11, 12 «Порядку присудження наукових ступенів», затвердженого постановою Кабінету Міністрів України від 24 липня 2013 р. № 567 (зі змінами від 30.12.2015 та 27.07.2016 р.) що висуваються до кандидатських дисертацій, а її авторка - Барбарук Л.В., заслуговує присудження наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 - інформаційні технології.

Офіційний опонент,
доцент кафедри спеціалізованих
комп'ютерних систем

Українського державного університету
залізничного транспорту,
доктор технічних наук, доцент



заповітну _____ 20 ____ р.
заявлену до розгляду
заявлену до розгляду
канцелярією
УкрДУЗТ

С.І. Доценко

С. І. Доценко
Олеся Алексєєва НМ