

**РОЗРОБКА ПАРАДИГМИ МЕТОДУ АНАЛІЗУ НАУКОВОЇ НОВИЗНИ ДОСЛІДЖЕННЯ
ЧЕРЕЗ МЕРЕЖІ МЕТАДАНИХ****DEVELOPMENT OF A PARADIGM OF THE METHOD OF ANALYSIS OF SCIENTIFIC
NOVELTY OF RESEARCH THROUGH METADATA NETWORKS**

**Прокопович-Ткаченко Д.І., к.т.н., доцент,
завідувач кафедри кібербезпеки та інформаційних технологій
Університет митної справи та фінансів**

**Саричев В.І., д.е.н., доцент,
професор кафедри економіки та економічної безпеки
Університет митної справи та фінансів**

**Костенко О.В., д.філос. (Ph.D.) з юр.н., старший дослідник,
завідувач наукової лабораторії теорії цифрової трансформації і права
Наукового центру цифрової трансформації і права
Державна наукова установа «Інститут інформації, безпеки і права
Національної академії правових наук України»**

**Соколенко О.І., аспірантка кафедри кібербезпеки
та інформаційних технологій
Університет митної справи та фінансів**

Стаття присвячена висвітленню проблем розробки парадигми методу аналізу наукової новизни дослідження на основі метаданих – MetaNovelNet. Метод поєднує використання семантичного аналізу, обробки природної мови (NLP) та графових моделей для виявлення унікальних зв'язків між основним змістом публікацій на визначену тему. Основна увага приділяється систематизації метаданих, таких як ключові слова, анотації, категорії, цитування, терміни публікації та міждисциплінарні зв'язки. Це дозволяє автоматизувати оцінку новизни дослідження, виявляти прогалини у його змісті та сприяти розвитку міждисциплінарних напрямків.

MetaNovelNet створює можливість аналізувати метадані через графові мережі, де вузли представляють наукові статті, а зв'язки відображають їх семантичну схожість або взаємозв'язки цитування. Використання методів машинного навчання, таких як кластеризація та виявлення аномалій, дозволяє ідентифікувати унікальні патерни, які свідчать про новизну. Алгоритми, такі як TF-IDF та нейронні мережі на базі трансформерів (наприклад, BERT), забезпечують глибокий аналіз текстів та об'єктивне порівняння публікацій.

Метод був протестований на базах даних Scopus та Web of Science, що підтвердило його ефективність у визначенні нових дослідницьких напрямків. Результати показали, що MetaNovelNet дозволяє зменшити ймовірність дублювання робіт, виявляти недостатньо розвинені теми та забезпечувати підґрунтя для наступних інноваційних ідей. Незважаючи на перспективність методу, існують виклики, пов'язані з неповнотою або недостатньою стандартизацією метаданих, а також складністю інтерпретації результатів. Подальші дослідження мають зосередитися на вдосконаленні алгоритмів, які враховують як об'єктивні, так і суб'єктивні аспекти новизни, а також інтеграції з іншими науковими системами для розширення аналізу.

Системи машинного навчання дедалі частіше використовуються для аналізу метаданих з метою визначення патернів, які вже свідчать про їх новизну. Наприклад, порівняння семантичних полів ключових слів і перевірка унікальності комбінаторики понять можуть визначити рівень оригінальності даної роботи. Метадані також дозволяють оцінювати новизну за допомогою аналізу цитування та джерел, на які посиляється дослідження, а також в результаті визначення прогалин у змісті вже існуючих робіт.

Застосування таких методів дозволяє створити інструменти для автоматизованої оцінки новизни, які сприяють зниженню ймовірності дублювання досліджень та посилюють міждисциплінарні зв'язки, і цим самим створюють можливості для нових відкриттів. Метадані, завдяки своїй структурованості, виступають ключовим ресурсом для розвитку таких інструментів, адже вони забезпечують прозорість і точність аналізу. Однак, існують виклики, такі як недостатність стандартизації метаданих або складність інтерпретації результатів алгоритмів, які все ще залишають простір для вдосконалення. Майбутні дослідження можуть зосереджуватися на розвитку більш адаптивних алгоритмів, які враховують не тільки об'єктивні, але й суб'єктивні аспекти новизни.

Ключові слова: наукова новизна, метадані, семантичний аналіз, обробка природної мови (NLP), графові моделі, машинне навчання, кластеризація, TF-IDF, нейронні мережі, трансформери, автоматизована оцінка, взаємозв'язки цитування, інноваційні ідеї, міждисциплінарний аналіз.

The article presents a new method for analyzing scientific novelty based on metadata – MetaNovelNet. The method combines semantic analysis, natural language processing (NLP), and graph models to identify unique connections between publications. The focus is on the systematization of metadata, such as keywords, annotations, categories, citations, and interdisciplinary links. This allows automating novelty assessment, identifying research gaps, and promoting interdisciplinary directions. MetaNovelNet analyzes metadata using graph networks, where nodes represent scientific articles and edges reflect their semantic similarity or citation relationships. The application of machine learning methods, such as clustering and anomaly detection, allows identifying unique patterns indicative of novelty. Algorithms like TF-IDF and transformer-based neural networks (e.g., BERT) ensure deep text analysis and effective comparison of publications.

The method was tested on Scopus and Web of Science databases, confirming its effectiveness in identifying new research directions. Results demonstrated that MetaNovelNet reduces the likelihood of duplicate work, identifies underdeveloped topics, and provides a foundation for innovative ideas. Despite the method's promise, challenges remain, such as incomplete or insufficient metadata standardization and complexity in interpreting results. Further research should focus on improving algorithms that consider both objective and subjective aspects of novelty and integrating them with other scientific systems to expand analysis. The novelty of scientific research is a critically important aspect of academic activity, determining a study's contribution to existing knowledge. Using metadata allows systematic novelty assessment through the analysis of parameters like keywords, categories, citations, and publication time. Metadata form the basis for automated uniqueness detection, ensuring standardized descriptions of studies, which facilitates comparison and classification.

Machine learning systems are increasingly used to analyze metadata to identify patterns indicative of novelty. For instance, comparing semantic fields of keywords and checking the uniqueness of concept combinations can reveal how original a study is. Metadata also allow novelty evaluation through analyzing citations and sources referenced in the research and identifying gaps in existing work.

The application of such methods enables creating tools for automated novelty evaluation, reducing redundant studies, and enhancing interdisciplinary links, creating opportunities for new discoveries. Metadata's structured nature is a key resource for developing such tools, ensuring transparency and precision in analysis. However, challenges such as insufficient metadata standardization or interpretation complexities still leave room for improvement. Future studies should focus on developing more adaptive algorithms that account for both objective and subjective aspects of novelty.

Key words: Scientific novelty, metadata, semantic analysis, natural language processing (NLP), graph models, machine learning, clustering, TF-IDF, neural networks, transformers, automated assessment, citation relationships, innovative ideas, interdisciplinary analysis.

Аналіз наукових джерел свідчить про необхідність впровадження інноваційних підходів до об'єктивного оцінювання дослідження та визначення його наукової новизни. Робота "Learning and Neural Networks: An Overview" [1] акцентує увагу на фундаментальних принципах функціонування нейронних мереж, що створює основу для розуміння перспектив застосування цих технологій у різних галузях науки. Автори підкреслюють важливість вдосконалення аналітичних методів для роботи з великими обсягами даних, що актуально для оцінки наукової новизни в міждисциплінарних дослідженнях.

Поряд з цим, у статті "Machine Learning and Deep Neural Networks" [2] досліджується потенціал глибокого навчання для роботи з великими масивами даних. Автори зазначають, що сучасні алгоритми глибокого навчання мають значні можливості для ідентифікації складних патернів, однак потребують адаптації для оцінки новизни наукових досліджень. Важливість цього аспекту підтверджується й у роботі "Toward Audio Beehive Monitoring: Deep Learning vs. Standard Machine Learning in Classifying Beehive Audio Samples" [3], де аналізуються переваги глибокого навчання порівняно зі стандартними методами. Це свідчить про необхідність впровадження таких алгоритмів у сферу аналізу метаданих.

Разом з тим, дослідження "Understanding Neural Networks for Machine Learning Using Microsoft Neural Network Algorithm" [4] розкриває необхідність застосування спеціалізованих нейронних мереж для розв'язання конкретних задач, зокрема автоматизованого аналізу текстових і графових даних. Результати цієї роботи вказують на перспективність створення інструментів для оцінки новизни шляхом порівняння семантичних взаємозв'язків та аналізу цитування.

Робота "Extreme Learning Machine for Interval Neural Networks" [5] демонструє ефективність новітніх методів машинного навчання у розв'язанні складних наукових задач. Ці методи, на думку авторів, можуть бути адаптовані для побудови моделей, які враховують як об'єктивні, так і суб'єктивні аспекти оцінки новизни. Аналогічні висновки представлені й у статті "Artificial Neural Networks in Medicine" [6], де обґрунтовується застосування інтелектуальних алгоритмів для аналізу міждисциплінарних зв'язків.

Принагідно зауважимо, що у схожій за спрямованістю роботі "Interpreting and Stabilizing Machine-Learning Parametrizations of Convection" [7], автори акцентують увагу на необхідності вдосконалення підходів до інтерпретації результатів алгоритмів машинного навчання. Вони підкреслюють важливість інтеграції інноваційних алгоритмів із існуючими системами для підвищення точності оцінки наукової новизни.

Таким чином, аналіз джерел свідчить про необхідність розробки та впровадження адаптивних інструментів, які б були здатні працювати з багатовимірними даними та сприяти виявленню нових наукових дослідницьких напрямів.

Тому метою публікації стала розробка, впровадження та тестування інноваційного методу аналізу наукової новизни дослідження на основі метаданих – MetaNovelNet. Основним завданням є створення автоматизованого інструменту, який поєднує семантичний аналіз, обробку природної мови (NLP) та графові моделі для ідентифікації унікальних зв'язків між публікаціями. Також завданням публікації є підтвердження ефективності запропонованого методу щодо виявлення прогалин у змісті досліджень, сприяння міждисциплінарним зв'язкам та мінімізації дублювання ідей і робіт у науковому середовищі.

Методологічно у сучасному науковому середовищі аналіз новизни є критично важливим для забезпечення унікальності досліджень, мінімізації дублювання та сприяння інноваціям. Зростання обсягу публікаційних даних ускладнює традиційні методи оцінки новизни, що в свою чергу вимагає розробки автоматизованих інструментів. Метадані, які включають ключові слова, категорії, цитування та інші структуровані дані, надають можливість проводити більш точний аналіз через їх стандартизованість. Отже, MetaNovelNet – метод, який запропонований у дослідженні, спрямований на вирішення цих викликів шляхом використання передових алгоритмів машинного навчання, таких як трансформери, та графових моделей.

Методологія розробки MetaNovelNet базується на поєднанні кількох технологічних підходів для аналізу наукової новизни дослідження. По-перше, для систематизації метаданих використано стандартні джерела, такі як бази даних Scopus та Web of Science. Ці метадані включають ключові слова, анотації, категорії, цитати та міждисциплінарні зв'язки. По-друге, для аналізу використано графові моделі, де вузли утворюються науковими публікаціями, а зв'язки – семантичними взаємозв'язками між ними або аналітикою цитування. Такий підхід дозволяє візуалізувати наукові зв'язки та ідентифікувати прогалини у вже існуючих дослідженнях.

Та, по-третє, в основу алгоритмічного аналізу покладено обробку природної мови. Для цього застосовано моделі, такі як TF-IDF для виявлення унікальних ключових слів, а також трансформери, які забезпечують семантичний аналіз текстів публікацій. Додатково в методології використано методи кластеризації та виявлення аномалій для ідентифікації унікальних патернів, а алгоритми машинного навчання дозволяють порівнювати публікації, визначати ступінь їх схожості та оцінювати новизну запропонованих концепцій.

Результати методу MetaNovelNet були перевірені на практиці шляхом тестування на обраних наборах даних. Це дозволило оцінити ефективність методу у виявленні нових напрямів досліджень, прогалин у змісті та потенційних можливостей для міждисциплінарного аналізу.

Алгоритм MetaNovelNet представлений як поетапний процес аналізу наукової новизни через обробку та аналіз метаданих публікацій. Основна мета алгоритму – ідентифікувати нові дослідницькі напрями та виявляти недостатньо досліджені теми шляхом автоматизованої обробки структурованої інформації. Використовуючи метадані, алгоритм аналізує зв'язки між науковими статтями, їх семантичну схожість, коректність цитування та відповідність патернів, виявляє аномалії й генерує звіти на основі отриманих даних.

Схематично алгоритм починається із завантаження метаданих із наукових баз даних, таких як Scopus чи Web of Science. На другому етапі дані очищуються, структуруються та готуються до подальшого аналізу (рис. 1).

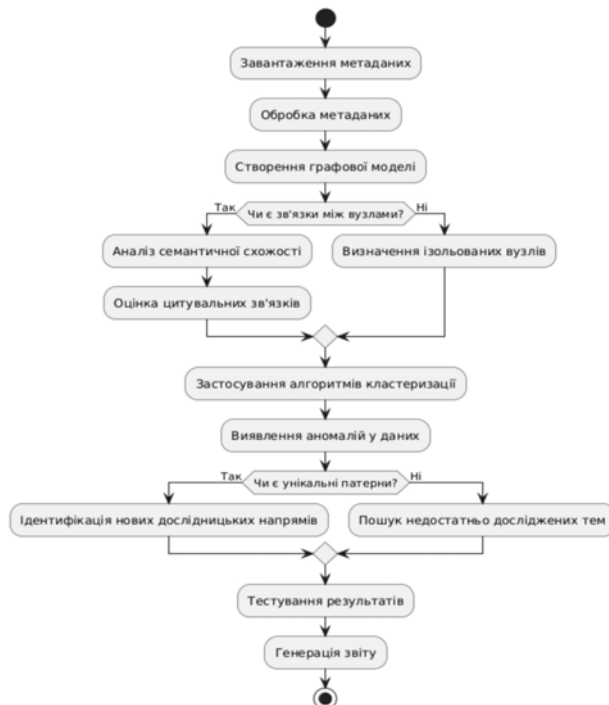


Рис. 1. Алгоритм застосування MetaNovelNet

Джерело: створено авторами.

Далі метадані перетворюються на графову модель, де вузли представлені наявними публікаціями, а зв'язки

між ними відображаються семантичними зв'язками або даними аналізу цитування.

Якщо між вузлами є стійкі зв'язки, здійснюється аналіз семантичної схожості та оцінка зв'язків цитування. Якщо зв'язків немає, визначаються ізольовані вузли, що можуть вказувати на унікальні або недостатньо досліджені теми. Після цього дані групуються в тематичні кластери за допомогою алгоритмів кластеризації для подальшого аналізу.

Далі відбувається ідентифікація незвичних або нових патернів у дослідницьких даних шляхом виявлення аномалій. Якщо виявлено унікальні патерни, визначаються нові дослідницькі напрями, а в разі їхньої відсутності здійснюється пошук недостатньо досліджених тем. На наступному етапі результати проходять перевірку на достовірність і точність. Завершується алгоритм створенням звіту, що містить висновки та рекомендації щодо перспективних дослідницьких напрямів або прогалин у знаннях.

Цей алгоритм дозволяє інтегрувати сучасні методи аналізу даних для автоматизованої оцінки наукової новизни дослідження (табл. 1).

Для демонстрації ефективності алгоритму MetaNovelNet було обрано випадковий набір даних, який пов'язаний із сучасними екологічними дослідженнями, зокрема темою біопластиків. У цьому прикладі використано метадані, отримані з провідних наукових баз даних, таких як Scopus, Web of Science і PubMed. Ці дані включали статті, що стосуються біопластиків, відновлюваної сировини та екологічних матеріалів. Метою аналізу було не лише ідентифікувати основні напрямки досліджень, але й виявити недостатньо досліджені тематики, які можуть стати основою для подальших інновацій у цій галузі.

Застосування запропонованого алгоритму було виконано шляхом проходження кількох ключових етапів, зокрема: обробки метаданих, створення графової моделі, семантичного аналізу та кластеризації. Варта уваги в дослідженні оцінка новизни в результаті аналізу частоти ключових слів і їхньої унікальності та виявлення аномальних патернів і даних.

Як результат, алгоритм дозволив не лише структурувати існуючі дослідження, але й запропонувати новий науковий напрямок. Дані табл. 1 наочно демонструють цей процес, узагальнюють результати аналізу та демонструють приклад пошуку інноваційного напрямку у науці за допомогою автоматизованого аналізу метаданих.

На радіальній діаграмі визначено відсотковий розподіл джерел за основними напрямками дослідження біопластиків. Візуалізація відображає найбільшу увагу, приділену темі біопластиків з кукурудзяного крохмалю (40%), а також інші напрямки, такі як полімери з водоростей (25%), сировина з лушпиння бананів (15%) та інші теми (20%) (рис. 2). Такий аналіз дозволяє ідентифікувати

провідні дослідницькі області та потенційні прогалини з позиції подальшого наукового пошуку.

Розподіл наукових джерел за напрямками досліджень є одним із ключових етапів для аналізу тенденцій, визначення актуальності тематик та виявлення прогалин у сучасній науці. Такий підхід дозволяє не лише оцінити, які напрями користуються найбільшою популярністю серед науковців, але й визначити теми, які потребують подальшого вивчення або удосконалення. Аналіз тенденцій у дослідженнях відіграє важливу роль у формуванні наукових стратегій, плануванні інноваційних проєктів та оптимізації розподілу ресурсів для підтримки перспективних ідей.

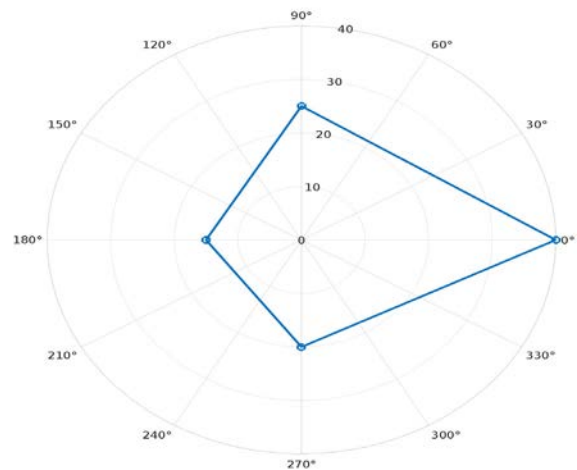


Рис. 2. Результати визначення розподілу джерел за основними напрямками дослідження, %

Джерело: створено авторами.

У даному прикладі розглянуто дослідження, що присвячені біопластам, які є однією з найактуальніших тем екологічних матеріалів завдяки їхньому потенціалу у вирішенні проблеми пластикових відходів.

Фокус аналізу зосереджений на кількох основних напрямках: біопластики з кукурудзяного крохмалю, полімери з водоростей, використання сировини з лушпиння бананів та інші теми. Біопластики з кукурудзяного крохмалю отримали найбільшу увагу науковців завдяки доступності сировини та відносно простій технології виробництва. Полімери з водоростей розглядаються як перспективна відновлювана сировина завдяки їхній екологічності та біорозкладності, хоча потребують подальших досліджень щодо

Таблиця 1

Варіант застосування MetaNovelNet для аналізу наукового напрямку з вивчення біопластиків

Етап	Опис
Збір метаданих	Метадані статей із баз Scopus, Web of Science, PubMed за ключовими словами: біопластики, відновлювана сировина, екологічні матеріали.
Обробка метаданих	Видалено дублікати, нормалізовано ключові слова, уніфіковано категорії для подальшого аналізу.
Створення графової моделі	Вузли представляють статті, а зв'язки базуються на семантичній схожості ключових слів, цитуваннях або категоріях.
Семантичний аналіз	Використано BERT для аналізу текстів анотацій та ключових слів. Статті розділено за тематиками, наприклад: біопластики з кукурудзяного крохмалю, полімери з водоростей.
Кластеризація та виявлення аномалій	Застосовано DBSCAN для виділення кластерів і виявлення аномальних публікацій, що стосуються нетипової сировини (наприклад, лушпиння бананів).
Оцінка новизни	TF-IDF виявив недостатньо досліджену тематику біопластиків із водоростей і прогалини в дослідженні їх впливу на екосистему.
Результати	Запропоновано новий напрямок: дослідження комбінованих полімерів із водоростей для покращення механічних властивостей біопластиків.

Джерело: створено авторами.

їхніх фізико-хімічних властивостей. Використання сировини з лушпиння бананів, незважаючи на те що має меншу кількість досліджень, демонструє інтерес до нестандартних рішень, що ґрунтуються на вторинній переробці відходів.

Поряд з цим, аналіз новизни наукових досліджень за допомогою алгоритму MetaNovelNet демонструє значний потенціал щодо вдосконалення методів систематизації, оцінки та ідентифікації перспективних напрямів у науці. У статті доведено, що автоматизація аналізу метаданих дозволяє значно зменшити ризики дублювання робіт і покращити якість міждисциплінарних досліджень. Використання семантичного аналізу, графових моделей і алгоритмів кластеризації дає змогу виявляти також унікальні патерни, які можуть слугувати основою для створення інноваційних підходів.

Проте, слід зазначити, що існує низка викликів, які обмежують застосування методу MetaNovelNet. По-перше, неповнота або недостатня стандартизація метаданих залишається серйозною проблемою для проведення глибокого аналізу. Це вимагає розробки узгоджених стандартів збирання та обробки даних. По-друге, складність інтерпретації результатів алгоритмів машинного навчання може знижувати їхню ефективність при аналізі специфічних або міждисциплінарних тем. Це свідчить про необхідність інтеграції алгоритмів з іншими науковими системами для покращення точності аналізу.

Обраний напрям досліджень, що пов'язаний із біопластиками, також ілюструє важливість ідентифікації прогалин у знаннях. Наприклад, результати вказують на

недостатнє вивчення тематики полімерів із водоростей, незважаючи на їхній потенціал для створення екологічно чистих матеріалів. Такий підхід до аналізу сприяє виявленню перспективних тем, які можуть бути ключовими для подальших досліджень.

Висновки. Запропонований у статті метод MetaNovelNet є ефективним інструментом для автоматизованого аналізу новизни наукових досліджень. Алгоритм дозволяє структурувати інформацію, аналізувати зв'язки між публікаціями, виявляти недостатньо досліджені теми та ідентифікувати унікальні патерни, які можуть слугувати основою для формування інноваційних підходів у дослідженнях.

Проведений аналіз на прикладі біопластиків показав, що найбільше досліджень зосереджено на біопластиках з кукурудзяного крохмалю, тоді як полімери з водоростей і сировина з лушпиння бананів мають значний потенціал, але потребують подальшого вивчення. Це підтверджує ефективність MetaNovelNet для виявлення наукових прогалин і перспективних напрямків дослідження.

Подальші дослідження повинні бути спрямовані на вдосконалення алгоритмів, зокрема адаптацію до різних дисциплін і інтеграцію з іншими системами для покращення якості аналізу. Крім того, необхідно зосередитися на вирішенні проблем стандартизації метаданих, що є важливим кроком для забезпечення об'єктивності, точності й ефективності запропонованого методу.

ЛІТЕРАТУРА

1. "Learning and Neural Networks: An Overview." IAES International Journal of Artificial Intelligence (IJ-AI), vol. 6, no. 2, June 1, 2017, P. 66. DOI: 10.11591/ijai.v6.i2.pp66-73.
2. Eberhard, Matthias, and Hatem Alkadhi. "Machine Learning and Deep Neural Networks." Journal of Thoracic Imaging, vol. 35, May 2020, pp. 17–20. DOI: 10.1097/rti.0000000000000482.
3. Kulyukin, Vladimir, Sarbajit Mukherjee, and Prakhar Amlathe. "Toward Audio Beehive Monitoring: Deep Learning vs. Standard Machine Learning in Classifying Beehive Audio Samples." Applied Sciences, vol. 8, no. 9, September 6, 2018, pp. 1573. DOI: 10.3390/app8091573.
4. Ramprasad, Nagesh. "Understanding Neural Networks for Machine Learning Using Microsoft Neural Network Algorithm." International Journal of Computer Applications, vol. 150, no. 3, September 15, 2016, pp. 32–38. DOI: 10.5120/ijca2016911481.
5. Yang, Dakun, Zhengxue Li, and Wei Wu. "Extreme Learning Machine for Interval Neural Networks." Neural Computing and Applications, vol. 27, no. 1, December 5, 2013, pp. 3–8. DOI: 10.1007/s00521-013-1519-3.
6. Kocsis, Zoltan Tamas. "Artificial Neural Networks in Medicine." Acta Technica Jaurinensis, vol. 12, no. 2, April 26, 2019, pp. 117–29. DOI: 10.14513/actatechjaur.v12.n2.497.
7. Brenowitz, Noah D., Tom Beucler, Michael Pritchard, and Christopher S. Bretherton. "Interpreting and Stabilizing Machine-Learning Parametrizations of Convection." Journal of the Atmospheric Sciences, vol. 77, no. 12, December 2020, pp. 57–75. DOI: 10.1175/jas-d-20-0082.1.
8. Dong, Yue. "Higher Order Neural Networks and Neural Networks for Stream Learning." Thesis, Université d'Ottawa, 2017. DOI: <http://hdl.handle.net/10393/35731>.
9. Menke, Joshua E. "Improving Machine Learning Through Oracle Learning." Thesis, 2007. Access Online.
10. Kalchbrenner, Nal. "Encoder-Decoder Neural Networks." Thesis, University of Oxford, 2017. DOI: <http://ora.ox.ac.uk/objects/uuid:d56e48db-008b-4814-bd82-a5d612000de9>.
11. Chen, Hsinchun. "Machine Learning for Information Retrieval: Neural Networks, Symbolic Learning, and Genetic Algorithms." Wiley Periodicals, 1995. DOI: <http://hdl.handle.net/10150/106427>.