

<https://doi.org/10.31891/2219-9365-2024-80-49>

УДК 004.89

БЛЕЦЬКИЙ Богдан

Вінницький національний технічний університет

<https://orcid.org/0009-0004-3553-1298>

МОКІН Віталій

Вінницький національний технічний університет

<https://orcid.org/0000-0003-1946-0202>

ІНТЕЛЕКТУАЛЬНА ТЕХНОЛОГІЯ АНАЛІЗУ ТЕМПОРАЛЬНОЇ УПОРЯДКОВАНOSTІ ЕЛЕМЕНТІВ СТРУКТУРНО-ЛОГІЧНОЇ СХЕМИ ОСВІТНЬОЇ ПРОГРАМИ

У статті представлено інтелектуальну технологію, яка дає змогу автоматизовано аналізувати освітні компоненти освітньої програми вищої освіти й визначити логічну послідовність їхнього викладання на основі структурування ключових понять та процесів життєвого циклу. Технологія спирається на сучасні інтелектуальні методи оброблення природної мови (NLP), зокрема із залученням великих мовних моделей (LLM), що дає змогу ідентифікувати та класифікувати порядок дисципліни в освітньому процесі відповідно до їхнього змісту. Підхід складається з 4 етапів. Спершу виконується підготовка даних, що містить відбір освітніх компонентів, релевантних визначеним спеціальним компетентностям і програмним результатам навчання. Далі в описі тем робочих програм цих компонентів виявляються ключові слова, згруповані по ключових поняттях. На третьому етапі визначаються слова, які відображають дії та процеси, що стосуються ключових понять. На останньому етапі використовується передтренована класифікаційна інтелектуальна модель, що здатна визначити етап багатостадійного процесу викладання, до якого варто віднести дисципліну (компонент) в освітньому процесі. Експериментальне застосування розробленої технології для реальної бакалаврської освітньої програми з інформаційних систем та технологій засвідчило високу результативність методу: ідентифіковано 367 ключових слів, які згруповані по 30 ключових поняттях, що охоплюють різні аспекти навчальних курсів. Розроблено тренувальний датасет для поняття "програмний продукт", на якому за допомогою алгоритму градієнтного бустингу досягнуто точність класифікації (ассигасу) для тристадійного процесу (тріад) на рівні 85–87%. Створена передтренована модель була успішно використана для розроблення рекомендацій для освітньої програми, яка аналізувалась. Практична цінність полягає у можливості впровадження даної технології для аналізу структурно-логічних схем, прискорюючи аналіз та підвищуючи якість розроблення та періодичного оновлення освітніх програм. У перспективі планується розширити методи оброблення великих текстових корпусів, вдосконалити класифікаційні алгоритми й протестувати підхід на суміжних галузях знань, щоб сформуувати універсальний інструмент для систематизації навчального процесу.

Ключові слова: інтелектуальна технологія, оброблення природномовного тексту, освітня програма, структурно-логічна схема, великі мовні моделі, темпоральне впорядкування.

BILETSKYI Bohdan, MOKIN Vitalii

Vinnitsia National Technical University

INTELLIGENT TECHNOLOGY FOR ANALYSIS OF TEMPORAL ORDERING OF ELEMENTS IN THE STRUCTURAL-LOGICAL SCHEME OF AN EDUCATIONAL PROGRAM

The article presents an intelligent technology that enables automated analysis of educational components within higher education curriculum and determines the logical sequence of their teaching based on the structuring of key concepts and life cycle processes. The technology relies on modern natural language processing (NLP) methods, particularly involving Large Language Models (LLM), which allows for identification and classification of course order in the educational process according to their content. The approach consists of 4 stages. Initially, data preparation is performed, which includes the selection of educational components relevant to defined specific competencies and learning outcomes. Next, keywords grouped by key concepts are identified in the description of topics from the syllabi of these components. In the third stage, words that reflect actions and processes related to key concepts are determined. At the final stage, a pre-trained classification AI model is used, capable of determining the stage of a multi-stage teaching process to which the discipline (component) should be assigned in the educational process. Experimental application of the developed technology for an actual bachelor's degree program in Information Systems and Technologies demonstrated high method effectiveness: 367 keywords were identified, grouped into 30 key concepts covering various aspects of the courses. A training dataset was developed for the concept of "software product," on which classification accuracy for a three-stage process (triads) of 85–87% was achieved using the gradient boosting algorithm. The created pre-trained model was successfully used to develop recommendations for the analyzed educational program. The practical value lies in the possibility of implementing this technology for analyzing structural and logical frameworks, accelerating analysis and improving the quality of development and periodic updates of educational programs. Future prospects include expanding methods for processing large text corpora, improving classification algorithms, and testing the approach in related fields of knowledge to form a universal tool for systematizing the learning process.

Keywords: intelligent technology, natural language processing, educational program, structural and logical framework, large language models, temporal ordering.

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Відповідно до чинних вимог щодо реалізації освітнього процесу слід регулярно переглядати та оновлювати структурно-логічну схему освітніх програм та зміст її обов'язкових освітніх компонентів, відповідно до рекомендацій стейкхолдерів. Але оновлення структурно-логічної схеми може внести протиріччя у зміст освітніх компонентів (дисциплін та ін.) і навпаки - зміни у темах занять робочих програм дисциплін призведуть до деякої невідповідності у структурно-логічній схемі програми. І якщо для освітніх програм рівня “магістратури” чи “доктор філософії” гарант і група забезпечення ще може це відслідковувати, то на рівні бакалаврату, де - багато дисциплін і їх викладають різні кафедри, це вже - нетривіальна задача. Зокрема, в інженерних та інформаційно-технологічних спеціальностях важливо дотримуватися логіки переходу від вивчення базових понять і принципів (наприклад, “теорія та моделювання”, “програмні підходи на рівні огляду і формування” тощо) до глибшого освоєння методів “розроблення”, “налаштування” й “впровадження”. Ще складнішими є триади, коли спочатку викладаються поняття, види, принципи функціонування та основи реалізації чогось (інформаційних систем, програмних продуктів, сервісів, технічних систем тощо), потім - як їх створювати та удосконалювати, а вже потім як використовувати на практиці для розв'язання прикладних задач та різні аспекти, які, при цьому, слід враховувати.

Звідси виникає потреба розробити автоматизовані технології аналізу змісту освітніх компонентів (ОК): силабусів і робочих програм дисциплін. Такий аналіз має ґрунтуватися на пошуку ключових слів і словосполучень (передусім, іменників із прикметниками), а також на їхньому поєднанні з типовими дієсловами, що вказують на конкретний тип дій у межах життєвого циклу: від “огляду й вивчення принципів” через “розроблення і створення” до “випробування та впровадження”. Результати такого аналізу дають змогу виявити, як саме в кожній дисципліні описано певний “етап” або “фазу” життєвого циклу розроблення програмного чи технічного продукту.

Подібний підхід дає змогу формувати рекомендації щодо узгодженості дисциплін за часовою віссю (ранжування тем, зв'язки між спорідненими поняттями, можливі дублювання). Це допоможе гарантувати, що студенти зможуть отримати потрібні базові знання (наприклад, “принципи програмування”) до початку курсу, де безпосередньо реалізуються “методи розроблення й впровадження програмного коду”. У контексті наукових і практичних завдань результати такого дослідження можуть стати основою для:

1. Модернізації навчальних планів, їхнього періодичного оновлення й вдосконалення відповідно до сучасних вимог ІТ-ринку.
2. Розроблення програмного інструменту, що допомагатиме викладачам і гарантам освітніх програм бачити міждисциплінарні зв'язки та впорядковувати навчальний контент.
3. Формування нових, більш обґрунтованих в науково-методичному плані підходів до укладання програм і силабусів.

Таким чином, проблема, в загальному вигляді, полягає у створенні інтелектуальної технології, що об'єднує засоби оброблення тексту природною мовою, а також — алгоритмічні процедури виявлення й класифікації ключових слів (іменників з прикметниками) та дієслів, які відображають певний етап навчання або стадію життєвого циклу. Важливо використати сучасні методи і підходи оброблення природномовного тексту (англ. “NLP”) з використанням великих мовних моделей (англ. “LLM”) та інші методи аналізу текстів. Практичне значення цієї технології полягає в тому, що автоматизований аналіз навчальних програм дозволяє підвищити узгодженість дисциплін (обов'язкових освітніх компонентів), оптимізувати їхній зміст та сприяє ефективнішому формуванню у студентів спеціальних компетентностей.

АНАЛІЗ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

Останніми роками застосування великих мовних моделей та технологій обробки природної мови в освітній сфері привернуло значну увагу. Дослідники вивчали кілька способів використання таких моделей для удосконалення навчальних програм, інтелектуального аналізу контенту, персоналізованого навчання та іншого. Повторюваною темою в літературі є зростаюча інтеграція LLM для підтримки освітніх цілей [1-3]. Наприклад, дослідники наголошують, що LLM можуть допомогти у створенні освітнього контенту, підвищити залученість студентів та персоналізувати навчальний досвід — хоча ці переваги супроводжуються потенційними недоліками, такими як крихкість моделі [1]. Існує консенсус, що для ефективного використання LLM освітяни повинні застосовувати стратегії перевірки вихідних даних моделі, керування процесом навчання та забезпечення відповідності цілям навчальної програми [1, 2].

Деякі дослідники вивчають системи автоматичного генерування вікторин та запитань. Їхньою метою є підвищення залученості студентів та адаптація навчального контенту до різноманітних здібностей студентів [5]. Тим часом, навчання з підкріпленням було використане для вдосконалення моделей генерування запитань, зменшуючи ризик погано структурованих запитань [3]. Хоча генерування вікторин стосується результатів навчання на мікрорівні.

Доповнюючи роботу над генеруванням запитань, дослідники також зосередилися на використанні LLM для проєктування навчання. Автоматичне створення повних навчальних планів, як це демонструють фреймворки на основі GPT-4 [4], або розроблення студентських планів для адаптивної підтримки в контекстах навчання на основі ігор [5], підкреслюють можливість систематичного охоплення структури та логіки освітніх компонентів. Графові подання знань, такі як ті, що запропоновані для персоналізації навчальних шляхів [6], також наголошують на важливості ідентифікації понять та процесів, які можуть бути пізніше відображені або послідовно розташовані відповідно до педагогічної логіки.

З точки зору оброблення природної мови (NLP), ідентифікація ключових понять та відповідних дій у текстових матеріалах була досліджена у відповідях на запитання у відкритій сфері [7], підходах передачі знань без навчання на нових даних [8, 9] та спеціалізованих для галузі застосунках NLI (Natural Language Inference) [10]. Ці дослідження підкреслюють корисні техніки для видобування термінології та відповідних процесів.

Синтезуючи ці напрямки досліджень — від адаптації великих мовних моделей, генерування запитань, редагування знань, персоналізації до методів навчання за навчальними програмами — посилюється ідея, що автоматизована система дійсно може ідентифікувати, класифікувати та послідовно розташовувати ключові поняття у численних навчальних програмах. Успіх такої системи залежить від надійних конвеєрів оброблення природної мови (включаючи складну інженерію запитів та групування по ключових поняттях), точного визначення дієслів для кожного етапу життєвого циклу та підходу до класифікації або ранжування, що розташовує кожен освітній компонент у правильній частині освітньої траєкторії.

Загалом, література підтверджує, що методи аналізу тексту на основі LLM, у поєднанні з передовими інструментами NLP та спеціалізованими знаннями галузі, можуть ефективно підтримувати проєктування та постійне вдосконалення освітніх програм. Інтегруючи встановлені техніки, такі як частотний аналіз, інженерія запитів, стає можливим розробити інтелектуальну технологію, яка може пропонувати удосконалення структурно-логічних схем та підвищувати узгодженість освітніх компонентів.

Важливо упорядковувати не стільки освітні компоненти освітніх програм в цілому, скільки — їх ключові теми, присвячені викладенню ключових понять. Особливістю такого упорядкування є те, що їх важливо упорядковувати не хронологічно — відносно конкретної часової осі (навіть, по семестрах — це, скоріше, задача гаранта та групи забезпечення, а — темпорально, тобто з'ясувати їх відносний порядок: який компонент передує іншому, а який — має викладатись після нього.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Мета дослідження полягає в розробленні інтелектуальної технології автоматизованого аналізу та темпорального упорядкування тем обов'язкових освітніх компонентів (ОК) освітньої програми, яка дає змогу:

- виявляти ключові поняття про результати роботи фахівців спеціальності (інформаційні чи технічні системи, технології, програмні продукти тощо) та — дієслова, пов'язані з різними етапами життєвого циклу розроблення і реалізації цих результатів роботи на практиці;
- визначати оптимальний темпоральний порядок компонентів-дисциплін на основі семантичного зв'язку між поняттями та діями, викладення яких передбачено у цих дисциплінах;

Підхід до досягнення мети дослідження пропонується реалізовувати у 4 етапи: підготовка даних, пошук ключових понять, пошук дієслів, що пов'язані з цими поняттями, хронологічне впорядкування дисциплін (рисунок 1).

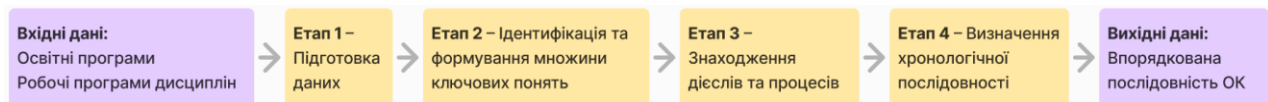


Рис. 1. Підхід до розроблення інтелектуальної технології автоматизованого аналізу та впорядкування змісту освітніх програм

ПІДГОТОВКА ДАНИХ

На підготовчому етапі головним завданням є знаходження потенційно подібних дисциплін та видобування з робочої програми (чи силабусу) опису тем змістових модулів і формування датасету, необхідного для подальшого аналізу й визначення темпоральних відносин між освітніми компонентами. Зокрема, передбачаються такі кроки:

- аналіз компетентностей і результатів навчання;
- визначення відповідних освітніх компонентів по таблицях відповідності освітньої програми;
- пошук і збирання навчальних матеріалів та збереження у датасеті.

Зокрема, спочатку експертним шляхом або за пошуком наявності певних слів чи частин слів важливих груп понять (наприклад “програма”, “система”, “продукт” тощо) знаходяться спеціальні

компетентності (СК) та результати навчання (РН), які містять ці слова. Також, для такого аналізу можна використати LLM. Результатом є номери цих СК та РН. Загалом, можна шукати і загальних компетентностях, але основні результати, як правило, мають бути саме в спеціальних.

На основі попередньо виділених СК та РН здійснюється відбір обов'язкових дисциплін, де формуються або розвиваються ці компетентності. Для цього використовується матриця відповідності компетентностей обов'язковим освітнім компонентам та матриця забезпечення програмних результатів навчання обов'язковими освітніми компонентами. Результатом цього є список ОК, які слід включити в датасет для подальшого аналізу й визначення темпоральних відносин між освітніми компонентами. З відповідних силабусів чи робочих програм цих ОК відбираються в датасет теми та їх описи, що містяться в розділах «Тематика дисципліни» чи «Змістові модулі».

В результаті маємо датасет, в якому містяться: назва ОК, список тем з описами тем.

ІДЕНТИФІКАЦІЯ ТА ФОРМУВАННЯ МНОЖИНИ КЛЮЧОВИХ ПОНЯТЬ

Після збирання вихідних даних про дисципліни, їхні теми та описи, виконується етап визначення ключових понять (напр. “програмний продукт”), які характеризуються різними ключовими словами в контекстах різних дисциплін (напр. “інформаційна система”, “програмне забезпечення”, “комп'ютерна програма” тощо). Тому для вилучення ключових понять передбачається 2 кроки:

- знаходження ключових слів для кожної ОК з датасету;
- групування ключових слів по ключових поняттях.

В результаті попереднього аналізу наявних даних, виявилося, що певні слова можуть бути частиною декількох словосполучень, які мають різний контекст. Наприклад, «Програма» може бути частиною різних змістовних словосполучень «Навчальна програма», «Прикладна програма», «Пакет програм». Відповідно було зроблено висновок, що ключове слово не достатньо характеризується одним словом, найчастіше для повноти сенсу використовується 2 слова (теоретично, їх може бути й більше), які мають подібні сполучення за частинами мови: «прикметник-іменник», чи «іменник-іменник».

Для виявлення ключових слів у потрібній формі доступні такі методи: застосування технологій оброблення природномовного тексту, метод аналізу частотності, чи за допомогою використання сучасних великих мовних моделей.

На основі попередньо виявлених ключових слів за допомогою експертного методу або із застосуванням великих мовних моделей чи гібридних методів обрані слова групуються по ключових поняттях (рисунок 2).

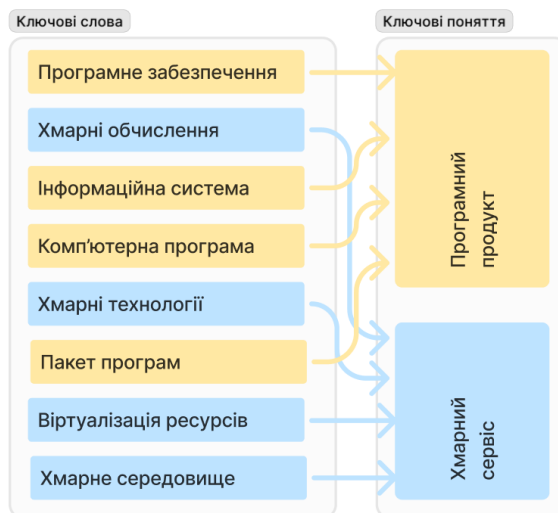


Рис. 2. Приклад співвідношення ключових слів до ключових понять

Для кожного ключового поняття визначається к-сть дисциплін в яких зустрічаються ключові слова, які характеризують ключове поняття. Відповідно до мети дослідження є сенс для подальшої роботи відібрати ті ключові поняття, які представлені в більше ніж одній дисципліні.

Результат даного етапу — є датасет ключових понять, що характеризуються відповідними групами ключових слів і представлені в більш ніж одній дисципліні.

ЗНАХОДЖЕННЯ ДІЄСЛІВ ТА ПРОЦЕСІВ

На цьому етапі зосереджується увага на виявленні дієслів чи слів-процесів, пов'язаних із ключовими поняттями, визначеними на попередньому етапі.

В описі теми дисципліни окрім ключових понять можна побачити і слова, які відображають дії з/над цими поняттями. Вони мають дві форми:

- дієслово, наприклад: “знаходити”, “обробляти”, “аналізувати”;
- віддієслівний іменник: “пошук”, “оброблення”, “аналіз”, що описують дію як процес.

Найчастіше в назвах та описах тем дисципліни в навчальних матеріалах використовується друга форма — віддієслівний іменник. Для коректності подальшого аналізу доречно обрати одну з форм і під час вилучення слів, перетворювати одну форму в іншу. Сучасні інструменти NLP або великі мовні моделі можуть допомогти в автоматичному перетворенні дієслів у віддієслівні іменники. Наприклад, бібліотеки Python, такі як SpaCy або NLTK [11].

Подальше узгодження «ключове поняття + дієслово/процес» уможливує встановлення послідовності етапів життєвого циклу для визначення порядку дисциплін, в яких вони викладаються.

Пропускаючи, що акредитовані програми містять в собі впорядковані дисципліни, зібравши їхні описи тем та проаналізувати порядок дисциплін певної тематики можна визначити які дієслова/процеси зустрічаються частіше на ранніх курсах, в середині освітнього процесу та - в кінці.

З використанням стандартних алгоритмів NLP можна визначити які дієслова є характерними виключно для певної стадії тріади.

Застосовуючи можливості LLM пошук та зіставлення ключових понять з пов’язаними словами-процесами можна виконувати одночасно. Однак, у деяких випадках мовна модель може пропонувати дієслова, що прямо не трапляються в тексті, але логічно впливають із контексту. Хоч вони формально й відсутні в описах дисциплін, проте здатні збагатити набір дій або уточнити їхній сенс.

В результаті маємо єдиний перелік дієслів з прив’язкою до ключового слова та відповідного ключового поняття.

ВИЗНАЧЕННЯ ТЕМПОРАЛЬНИХ ВІДНОСИН МІЖ ОСВІТНІМИ КОМПОНЕНТАМИ

На четвертому етапі виконується визначення темпоральних відносин дисциплін у межах освітніх компонентів, що базується на аналізі ключових понять і пов’язаних із ними процесів. Для цього розробляється розмічений датасет і класифікаційна модель, яка допомагає впорядковувати дисципліни за їхнім змістом і логічним зв’язком.

Спершу формування датасету починається зі збирання блок-схем акредитованих освітніх програм та аналізу навчальних матеріалів дисциплін, доступних у відкритих джерелах. Усі матеріали аналізуються з метою виявлення ключових слів і дієслів/процесів. Кожна дисципліна відповідно до своєї позиції в блок-схемі визначається як така, що належить до першої, другої або третьої частини тріади. Перша частина включає дисципліни, що викладають базові поняття і принципи функціонування найчастіше знаходяться на 1 курсі, друга – дисципліни, які зосереджуються на розробленні та вдосконаленні продуктів, такі дисципліни зазвичай знаходяться на 2-3 курсах, а третя – дисципліни, що демонструють практичне застосування цих продуктів і знаходяться на останніх курсах.

Кожна дисципліна з акредитованих програм обробляється таким самим способом: знаходяться ключові поняття, пов’язані з ними дієслова/процеси та визначається місце цієї дисципліни в тріаді. Відповідно, формується датасет, що містить назви дисциплін, ключові слова, дієслова і позицію дисципліни у тріаді. Цей датасет стає основою для навчання класифікаційної моделі, яка визначатиме порядок дисциплін на основі їхнього змісту. Якщо датасет - не достатньо великий, тоді його можна доповнити інформацією з документації чи іншої довідкової літератури, присвяченій життєвому циклу функціонування того об’єкту, який відповідає тому поняттю (наприклад, проектування, супровід та використання програмного забезпечення, створення і використання інформаційно-вимірювальних систем тощо).

Далі виконується навчання класифікаційної моделі. Після навчання модель перевіряється за метриками точності, повноти і F1-міри, а також тестується на нових дисциплінах, що не входили до датасету. Це дозволяє оцінити її здатність коректно визначати послідовність дисциплін і враховувати взаємозв’язки між ними. У результаті цього етапу створюється класифікаційна модель, яка дозволяє визначати порядок дисциплін у межах тріад. Вона забезпечує логічну узгодженість освітніх програм, підвищує якість їхньої структурної організації.

ЕКСПЕРИМЕНТ, РЕЗУЛЬТАТИ ТА ДИСКУСІЯ

Для експерименту використано освітні матеріали Вінницького національного технічного університету (ВНТУ), зокрема бакалаврську освітню програму “Прикладні інформаційні технології” зі спеціальності 126 «Інформаційні системи та технології». Проведено аналіз за ключовим поняттям «програмний продукт».

На першому етапі було проаналізовано наявні РН та СК, обрані з них ті, які є дотичними до ключового поняття. Використовуючи матриці відповідності, було визначено, які ОК вивчають необхідний напрямок. На основі навчальних матеріалів цих ОК сформовано датасет, який містив назви ОК та тексти, що описують тематику й зміст відповідних дисциплін (рис. 1).

За допомогою великої мовної моделі ChatGPT-4 та серії промптів до кожної ОК було сформовано списки ключових слів та здійснено їх об'єднання в загальний список з 376 елементів і - групування по 30 ключових поняттях.

З відкритих джерел зібрано корпус текстів, який складався з трьох частин, відповідно до тріадного поділу життєвого циклу програмного продукту. Застосовуючи метод TF-IDF [11], для кожної частини тріади визначено слова, що найчастіше зустрічаються у певній частині тріади й не трапляються в інших. Розподіл унікальних та спільних слів словника між стадіями тріади представлено на рисунку 3. Загальна кількість слів у словнику зібраного датасету - 839 слів. Для кожного ключового поняття, яке присутнє у більше ніж трьох дисциплінах, за допомогою серії запитів до LLM виявлено відповідні процеси та дієслова, які з ними пов'язані.



Рис. 3. Діаграма Венна розподілу унікальних та спільних слів між стадіями тріади

Сформований у такий спосіб датасет використано як тренувальний датасет для моделі класифікації отриманого набору даних щодо виявлення в якій частині навчальної тріади знаходиться дисципліна (рис. 3).

Скорочена Назва дисципліни №	Навчальні процеси														
	Аналіз	Використання	Забезпечення	Застосування	Ідентифікація	Керування	Моделювання	Модернізація	Налагодження	Оцінка	Проектування	Реалізація	Розроблення	...	Створення
АПЗ	1	1	0	0	1	1	1	1	1	1	1	1	0	...	1
ІВ	2	0	1	1	0	1	0	1	0	1	0	1	0	...	0
ТБД	3	1	1	1	1	0	0	0	1	1	0	1	0	...	1

Рис. 4. Фрагмент навчального датасету

Було побудовано та навчено багатоцільову модель класифікації на основі градієнтного бустингу для прогнозування трьох класів. Після навчання модель продемонструвала достатньо високу ефективність: точність (ассугасу) склала 85-87%, а значення F1-міри для кожної цільової змінної свідчили про збалансовану продуктивність (таблиця 1), хоча третю стадію ідентифікувати важче.

Таблиця 1.

Оцінки навчання моделі

Target	Precision	Recall	F1-score	Accuracy
Перша стадія	0,9	0,8	0,85	0,86
Друга стадія	0,91	0,87	0,87	0,87
Третя стадія	0,78	0,85	0,86	0,85

В результаті тестування цієї моделі сформована рекомендація щодо порядку викладання в межах освітньої програми дисциплін «Алгоритмізація та програмування», «Об'єктно-орієнтоване програмування» та «Управління ІТ-проектами», які містять ключове поняття «програмний продукт», (рисунок 5).

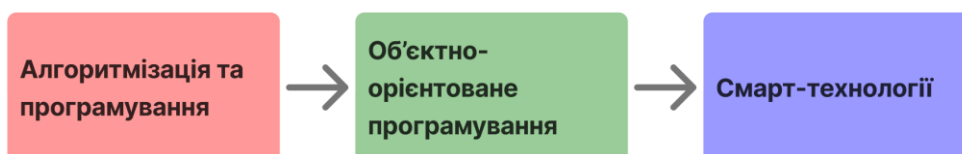


Рис. 5. Запропонована послідовність викладання дисциплін в межах освітньої програми

Оскільки в зібраному датасеті було виявлено 367 ключових слів, які об'єднуються в 14 ключових поняття, є сенс провести подібне дослідження і для інших ключових поняття. Це, своєю чергою, дозволить надати значний обсяг рекомендацій для можливих змін у структурно-логічну схему.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

Розроблено інтелектуальну технологію автоматизованого аналізу темпорального упорядкування тем освітніх компонентів. Створена технологія успішно вирішує всі поставлені завдання: забезпечує виявлення ключових понять (367 слів об'єднані в 30 ключових понять) та пов'язаних з ними дієслів життєвого циклу; дозволяє визначати оптимальний темпоральний порядок дисциплін з високою точністю класифікації (85-87%); на основі якого можна створити рекомендації щодо узгодження навчального контенту, що підтверджено на прикладі аналізу дисциплін 126 спеціальності рівня “бакалавр” за освітньою програмою “Прикладні інформаційні технології” ВНТУ, пов'язаних з поняттям “програмний продукт”.

В роботі виявлено та обґрунтовано актуальну потребу в автоматизації аналізу змісту освітніх компонентів для забезпечення логічної послідовності їх викладання. Показано, що використання сучасних методів оброблення природної мови та великих мовних моделей є ефективним підходом до розв'язання даної задачі.

Проведений аналіз досліджень та публікацій підтвердив доцільність використання NLP технологій, зокрема LLM, для аналізу освітнього контенту. Виявлені успішні практики застосування мовних моделей для проектування навчання та аналізу освітніх матеріалів демонструють перспективність цього напрямку. Встановлено, що найбільш ефективним є комбінування методів частотного аналізу та інженерії запитів для великих мовних моделей при роботі з освітнім контентом.

У методологічному аспекті запропоновано комплексний чотириетапний підхід до аналізу та темпорального впорядкування змісту освітніх програм. Розроблено методику виявлення та групування ключових понять на основі аналізу освітніх компонентів, а також створено алгоритм визначення темпоральних відносин дисциплін на основі аналізу процесів життєвого циклу. Такий підхід забезпечує системність та повноту аналізу освітніх програм.

Сформовано тренувальний датасет, який дозволив натренувати інтелектуальну модель, яка з точністю від 85% до 87% за метриками F1 та Accurasy визначає віднесення опису теми до кожної зі стадій тріади життєвого циклу опису основ, створення та використання програмного забезпечення. Експериментальне дослідження цієї моделі, проведене на прикладі освітньої програми “Прикладні інформаційні технології” спеціальності 126, підтвердило ефективність розробленої методології. Виявлення 367 ключових слів, які об'єднуються в 30 ключових понять, демонструє масштабованість методології.

Практична значущість роботи полягає в тому, що розроблена технологія дозволяє автоматизувати процес аналізу та оптимізації структурно-логічних схем освітніх програм. Створений інструментарій може бути ефективно використаний для підвищення якості освітніх програм та їх відповідності сучасним вимогам.

Подальші дослідження доцільно спрямувати на розширення аналізу на інші освітні програми, удосконалення класифікаційної моделі шляхом збільшення навчального датасету, розроблення програмного інструментарію для автоматизації усіх етапів запропонованого підходу (експертний підхід на окремих стадіях можна спробувати замінити на LLM) та дослідження можливостей застосування методології для інших спеціальностей і галузей. Отримані результати створюють надійну основу для подальшого розвитку автоматизованих систем аналізу та оптимізації освітніх програм.

Література

1. Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., та ін. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education // *Learning and Individual Differences*. – 2023. – Т. 103, Article ID: 102274. – Режим доступу: <https://doi.org/10.1016/j.lindif.2023.102274>.
2. Sreekanth D., Dehbozorgi N. Enhancing Engineering Education Through LLM-Driven Adaptive Quiz Generation // C-DAY FALL 2023: Proceedings of the Conference. Kennesaw, 2023. P. 45–52. URL: <https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1399&context=cday>
3. Fahl W. GraphWiseLearn: Personalized Learning through Semantified TEL, Leveraging QA-Enhanced LLM-Generated Content [Електронний ресурс]. – Режим доступу: <https://2024.eswc-conferences.org/wp-content/uploads/2024/05/77770405.pdf>.
4. Hu B., Zheng L., Zhu J., Ding L., Wang Y., Gu X. Teaching Plan Generation and Evaluation with GPT-4: Unleashing the Potential of LLM in Instructional Design // *IEEE Transactions on Learning Technologies*. – 2024. – Т. 17, С. 1471-1485. – Режим доступу: <https://doi.org/10.1109/tlt.2024.3384765>.

5. Goslen A., Kim Y.J., Rowe J., Lester J. LLM-Based Student Plan Generation for Adaptive Scaffolding in Game-Based Learning Environments // *International Journal of Artificial Intelligence in Education*. – 2024. – Режим доступу: <https://doi.org/10.1007/s40593-024-00421-1>.
6. Ma K., Cheng H., Zhang Y., Liu X., Nyberg E., Gao J. Chain-of-Skills: A Configurable Model for Open-Domain Question Answering // *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. – Торонто, 2023. – С. 1599-1618. – Режим доступу: <https://doi.org/10.18653/v1/2023.acl-long.89>.
7. Winata G., Xie L., Radhakrishnan K., Gao Y., Preotiuc-Pietro D. Efficient Zero-Shot Cross-Lingual Inference via Retrieval // *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*. – Nusa Dua, November 2023. – С. 93-104. – Режим доступу: <https://doi.org/10.18653/v1/2023.ijcnlp-short.11>.
8. Conceição S.I.R., Sousa F., Silvestre P., Couto F.M. LasigeBioTM at SemEval-2023 Task 7: Improving Natural Language Inference Baseline Systems with Domain Ontologies // *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. – Торонто, 13-14 липня 2023. – С. 10-15. – Режим доступу: <https://doi.org/10.18653/v1/2023.semeval-1.2>.
9. Kolagar Z., Zarcone A. HumSum: A Personalized Lecture Summarization Tool for Humanities Students Using LLMs // *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*. – St. Julians, березень 2024. – С. 36-70. – Режим доступу: <https://aclanthology.org/2024.personalize-1.4>.
10. Gong Z., Liu J., Wang Q., Yang Y., Wang J., Wu W., та ін. PreQuant: A Task-Agnostic Quantization Approach for Pre-Trained Language Models // *Findings of the Association for Computational Linguistics: ACL 2023*. – Toronto, 9-14 July 2023. – С. 8065-8079. – Режим доступу: <https://doi.org/10.18653/v1/2023.findings-acl.511>.
11. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник / В. Б. Мокін, М. В. Дратований – Вінниця : ВНТУ, 2024. – 258 с.

References

1. Kasneci E., Sessler K., Küchemann S., Bannert M., Dementieva D., Fischer F., та ін. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education // *Learning and Individual Differences*. – 2023. – Т. 103, Article ID: 102274. – Access mode: <https://doi.org/10.1016/j.lindif.2023.102274>
2. Sreekanth D., Dehbozorgi N. Enhancing Engineering Education Through LLM-Driven Adaptive Quiz Generation // *C-DAY FALL 2023: Proceedings of the Conference*. Kennesaw, 2023. P. 45–52. URL: <https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1399&context=cday>
3. Fahl W. GraphWiseLearn: Personalized Learning through Semantified TEL, Leveraging QA-Enhanced LLM-Generated Content [Electronic resource]. – Access mode: <https://2024.eswc-conferences.org/wp-content/uploads/2024/05/77770405.pdf>
4. Hu B., Zheng L., Zhu J., Ding L., Wang Y., Gu X. Teaching Plan Generation and Evaluation with GPT-4: Unleashing the Potential of LLM in Instructional Design // *IEEE Transactions on Learning Technologies*. – 2024. – Т. 17, С. 1471-1485. – Access mode: <https://doi.org/10.1109/tlt.2024.3384765>
5. Goslen A., Kim Y.J., Rowe J., Lester J. LLM-Based Student Plan Generation for Adaptive Scaffolding in Game-Based Learning Environments // *International Journal of Artificial Intelligence in Education*. – 2024. – Access mode: <https://doi.org/10.1007/s40593-024-00421-1>
6. Ma K., Cheng H., Zhang Y., Liu X., Nyberg E., Gao J. Chain-of-Skills: A Configurable Model for Open-Domain Question Answering // *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. – Toronto, 2023. – С. 1599-1618. – Access mode: <https://doi.org/10.18653/v1/2023.acl-long.89>
7. Winata G., Xie L., Radhakrishnan K., Gao Y., Preotiuc-Pietro D. Efficient Zero-Shot Cross-Lingual Inference via Retrieval // *Proceedings of the 13th International Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 2: Short Papers)*. – Nusa Dua, November 2023. – С. 93-104. – Access mode: <https://doi.org/10.18653/v1/2023.ijcnlp-short.11>
8. Conceição S.I.R., Sousa F., Silvestre P., Couto F.M. LasigeBioTM at SemEval-2023 Task 7: Improving Natural Language Inference Baseline Systems with Domain Ontologies // *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. – Торонто, 13-14 липня 2023. – С. 10-15. – Access mode: <https://doi.org/10.18653/v1/2023.semeval-1.2>
9. Kolagar Z., Zarcone A. HumSum: A Personalized Lecture Summarization Tool for Humanities Students Using LLMs // *Proceedings of the 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE 2024)*. – St. Julians, March 2024. – С. 36-70. – Access mode: <https://aclanthology.org/2024.personalize-1.4>
10. Gong Z., Liu J., Wang Q., Yang Y., Wang J., Wu W., та ін. PreQuant: A Task-Agnostic Quantization Approach for Pre-Trained Language Models // *Findings of the Association for Computational Linguistics: ACL 2023*. – Toronto, July 9-14, 2023. – С. 8065-8079. – Access mode: <https://doi.org/10.18653/v1/2023.findings-acl.511>
11. Data Science: Machine Learning and Data Mining: an Electronic Textbook / V. B. Mokin, M. V. Dratovanyi - Vinnytsia: VNTU, 2024. 258 p.