

УДОСКОНАЛЕНИЙ АЛГОРИТМ МУЛЬТИМОДАЛЬНОЇ ОБРОБКИ ДАНИХ В ІНТЕЛЕКТУАЛЬНОМУ КУЛІНАРНОМУ ПОМІЧНИКУ

Вінницький національний технічний університет

Анотація

Удосконалено алгоритм мультимодальної обробки кулінарної інформації для формування персоналізованих рекомендацій, який поєднує аналіз текстових запитів, розпізнавання фотографій інгредієнтів та інтеграцію векторних репрезентацій у спільному просторі ознак.

Ключові слова: кулінарні дані, мультимодальна інтеграція, GPT, EfficientNet, CLIP, генерація рецептів.

Abstract

An improved algorithm of multimodal data processing for generating personalized cooking recommendations is developed. It integrates text query analysis, ingredient image recognition, and unified vector-space representations for accurate recipe generation.

Keywords: culinary data, multimodal integration, GPT, EfficientNet, CLIP, recipe generation.

Вступ

У сучасному світі обсяг цифрових кулінарних даних стрімко зростає, що ускладнює отримання точних та персоналізованих рекомендацій для користувачів. Традиційні пошукові системи, які базуються на простому збігу ключових слів, не здатні забезпечити високий рівень релевантності, оскільки не враховують контекст запиту, візуальні характеристики інгредієнтів та індивідуальні дієтичні обмеження користувачів. Це створює потребу в розробці інтелектуальних технологій, здатних працювати з даними різної природи — текстовими описами, фотографіями страв та інгредієнтів, метаданими рецептів та харчовими параметрами.

Удосконалення алгоритмів обробки таких даних передбачає відхід від класичних методів пошуку та перехід до підходів, заснованих на семантичному аналізі, мультимодальній інтеграції та глибинному розумінні змісту запиту. Саме семантичний аналіз дозволяє системі інтерпретувати не окремі слова, а взаємозв'язки між інгредієнтами, способами приготування та дієтичними вимогами [1]. Додатково сучасні моделі комп'ютерного зору дають змогу здійснювати автоматичне розпізнавання інгредієнтів на фотографіях, що значно розширює можливості користувача при формуванні запиту.

Таким чином, побудова інтелектуального модуля, який використовує семантичний пошук, аналіз зображень та мультимодальну інтеграцію даних, є перспективним напрямом створення точних та гнучких систем кулінарних рекомендацій.

Результати досліджень

У сучасних інтелектуальних кулінарних системах застосовуються різні підходи до обробки вхідних даних, які формуються користувачем у вигляді текстових описів, переліків інгредієнтів, дієтичних обмежень або зображень продуктів. Кожен підхід вирішує окремий клас задач та має свої переваги й обмеження. Їх аналіз дозволяє виділити ті компоненти, які доцільно інтегрувати у вдосконалений алгоритм мультимодальної обробки даних:

1. Обробка даних на основі ключових слів:

Цей підхід передбачає виділення та інтерпретацію ключових термінів, що описують інгредієнти, технологічні етапи або характеристики страви. Він є одним із найпростіших способів первинної обробки інформації та використовується для класифікації та попереднього структурування даних.

Переваги цього методу полягають у його швидкодії та мінімальних обчислювальних затратах. Однак він має суттєве обмеження — відсутність урахування контексту. Система не здатна повною

мірою розпізнати приховані наміри користувача чи коректно інтерпретувати складні або нечіткі формулювання. У випадках, коли користувач описує страву загальними словами або використовує синонімічні назви інгредієнтів, точність такого підходу значно знижується.

2. Персоналізований пошук:

Персоналізований пошук передбачає адаптацію результатів під конкретного користувача, враховуючи його дієтичні потреби, попередні вибори, історію взаємодій, популярні страви та індивідуальні обмеження. Такий підхід широко застосовується в кулінарних сервісах, оскільки харчові вподобання значною мірою залежать від індивідуального стилю життя, здоров'я та культурного контексту.

Алгоритми персоналізованого пошуку використовують статистичні моделі та методи машинного навчання для виділення закономірностей у виборі користувача. Це дозволяє формувати рекомендації з вищою релевантністю, проте створює ризик «фільтр-бульбашки», коли користувачу пропонуються лише подібні страви, а різноманітність поступово зменшується. Крім того, для повноцінного функціонування підхід потребує накопичення історії та активного зворотного зв'язку, що обмежує його застосовність у системах, якими користуються нерегулярно або анонімно.

3. Семантична обробка даних:

Семантичний пошук спрямований на розуміння змісту та наміру запиту, а не лише на співставлення ключових слів [2]. У кулінарних системах цей підхід є надзвичайно важливим, оскільки користувачі часто вводять запити на зразок:

- «що приготувати з того, що є в холодильнику»,
- «щось легке з куркою»,
- «веганська вечеря без сої».

Традиційні алгоритми не здатні коректно інтерпретувати такі запити, однак трансформерні моделі (GPT, BERT-подібні) дозволяють визначати контекст, приховані наміри та зв'язки між інгредієнтами, техніками приготування та дієтичними вимогами [3].

У сучасних системах семантичний пошук часто реалізований через аналіз метаданих, однак це не завжди точно відображає справжній зміст рецепта. Повноцінне семантичне зіставлення потребує векторних репрезентацій текстів та використання глибинних моделей, що значно збільшує точність, але потребує більших обчислювальних ресурсів.

Завдяки високій здатності до інтерпретації текстових описів та контекстних зв'язків семантичний пошук є основою роботи сучасних інтелектуальних кулінарних систем. Його поєднання з аналізом зображень та мультимодальною інтеграцією формує основу високоточних рекомендаційних модулів.

Переваги та недоліки цих підходів наведено в таблиці 1.

Таблиця 1 – Переваги та недоліки підходів пошуку різноформатних даних в сучасних пошукових системах

Підхід	Переваги	Недоліки
Обробка даних на основі ключових слів	- Простий і поширений; - Ефективний та гнучкий;	- Може бути неточним; - Не враховує контекст запиту;
Персоналізований пошук	- Підвищує релевантність результатів - Популярний через зростання обсягів даних	- Ризик "фільтр-бульбашки" - Потребує зворотного зв'язку від користувача
Семантична обробка даних	- Краще розуміння запитів та контексту - Надає більш релевантні результати - Можливість використання потужних ресурсів для високої точності	- Повноцінна імплементація потребує значних ресурсів

На основі порівняльного аналізу існуючих підходів можна зробити висновок, що застосування семантичної обробки даних як базового механізму забезпечує суттєве підвищення точності формування підсумкових результатів. Завдяки перетворенню вхідних матеріалів у векторні представлення та використанню моделей, здатних працювати з різними форматами даних, система демонструє кращу узгодженість між інтерпретацією запиту та сформованими рекомендаціями. Це дає змогу підвищити загальну ефективність інформаційної технології та забезпечити стабільну якість

роботи незалежно від типу введеної інформації.

З урахуванням виявлених переваг семантичного підходу, удосконалений алгоритм обробки кулінарних даних в інтелектуальному помічнику для приготування страв має бути побудований на основі мультимодальної інтеграції різнотипних вхідних даних. Це передбачає підвищену точність створення релевантних рецептів та можливість формування персоналізованих інструкцій приготування на основі текстових запитів, зображень інгредієнтів або страв, а також дієтичних параметрів за рахунок використання єдиного стандарту запису інформації у вигляді векторної репрезентації.

Таким чином, удосконалений алгоритм роботи інтелектуального помічника для приготування страв складається з двох основних компонентів: алгоритму мультимодальної обробки кулінарних даних та алгоритму генерації рецептів. У межах даної роботи детально розглядається перша складова – алгоритм обробки різнотипних вхідних даних. Особливістю цього підходу є використання єдиного стандарту запису інформації у вигляді векторної репрезентації, що забезпечує підвищену точність створення релевантних рецептів та можливість формування персоналізованих інструкцій на основі текстових запитів, зображень інгредієнтів або страв, а також дієтичних параметрів.

Архітектура алгоритму обробки кулінарних даних:

Для обробки кулінарних даних потрібно використовувати алгоритми, які можуть призвести різні формати до одного формату. У випадку обробки зображень інгредієнтів та текстових запитів користувачів найдоцільніше буде використовувати алгоритми, що перетворюють дані на векторні репрезентації:

- для обробки текстових запитів користувачів використовуються алгоритми, побудовані на основі трансформерних архітектур типу GPT, BERT або RoBERTa;
- для аналізу зображень інгредієнтів застосовуються CNN-моделі типу ResNet або EfficientNet для розпізнавання продуктів [4];
- для поєднання візуальної та текстової інформації використовуються мультимодальні моделі типу CLIP.

Щоб отримати контекстуальну репрезентацію вхідних параметрів (текстового запиту, розпізнаних інгредієнтів, дієтичних обмежень), необхідно застосувати алгоритми, здатні враховувати контекст та зв'язки між різними характеристиками бажаної страви, а не лише окремі слова або параметри. Саме тому для вирішення цієї проблеми доцільно використати алгоритми типу GPT, BERT або трансформерні моделі з механізмами уваги.

У межах даної роботи доцільно застосувати саме трансформерні моделі типу GPT для обробки вхідних даних, оскільки цей алгоритм поєднує високу якість контекстуальних репрезентацій з можливістю врахування семантичних зв'язків між інгредієнтами, способами приготування та дієтичними характеристиками страв. Додатково він також надає можливість додаткового навчання, що в свою чергу дозволяє створити модель, яка буде справлятися з кулінарними задачами на високому рівні. Хоча він і потребує значних обчислювальних ресурсів, проте для завдань розуміння кулінарних запитів користувачів репрезентації на рівні контексту є необхідним.

Компоненти системи обробки візуальних даних:

Розглянемо основні алгоритмічні компоненти системи обробки візуальних кулінарних даних:

1. EfficientNet для розпізнавання інгредієнтів

Алгоритм типу CNN для розпізнавання зображень дозволяє перетворювати візуальний вміст фотографій інгредієнтів у текстову інформацію про наявні продукти для подальшої обробки. Архітектура EfficientNet заснована на поєднанні переваг балансованого масштабування глибини, ширини та роздільної здатності нейронної мережі [4]. Поєднуючи переваги цих параметрів, EfficientNet забезпечує високу точність розпізнавання харчових продуктів на зображеннях, водночас зберігаючи ефективність використання обчислювальних ресурсів.

Основні переваги EfficientNet:

- висока точність розпізнавання інгредієнтів на фотографіях;
- ефективне використання обчислювальних ресурсів;
- швидка робота системи розпізнавання;
- можливість обробки великих обсягів зображень харчових продуктів.

2. CLIP для мультимодальної інтеграції

Для забезпечення максимальної ефективності системи необхідно не лише розпізнати інгредієнти на зображенні, а й об'єднати цю візуальну інформацію з текстовими запитами користувача в єдиному векторному просторі. Саме для цього застосовуються мультимодальні моделі типу CLIP.

CLIP (Contrastive Language-Image Pre-training) – архітектура CLIP дозволяє створювати спільний векторний простір для зображень та тексту, що робить можливим безпосереднє порівняння візуальної інформації про інгредієнти з текстовими описами вподобань користувача. На відміну від традиційного підходу, де спочатку розпізнаються об'єкти на зображенні, а потім окремо обробляється текст, CLIP навчається розуміти відповідності між візуальними та текстовими концепціями одночасно. Це дозволяє системі природно інтерпретувати, наприклад, фотографію овочів разом із запитом "легка вечеря на двох" або "низькокалорійна страва", формуючи єдину контекстуальну репрезентацію для генерації відповідного рецепту (рис.1).

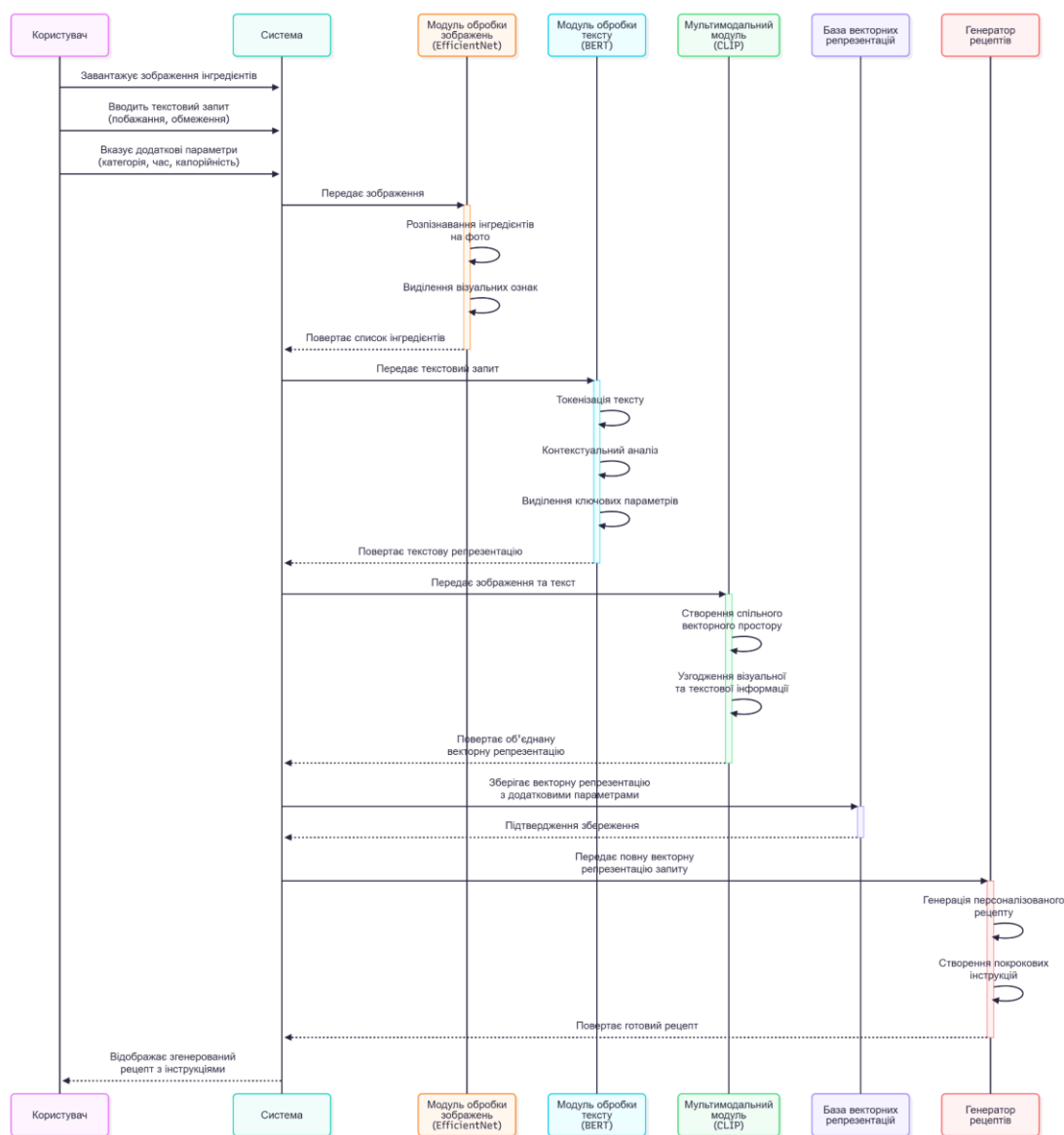


Рисунок 1 – UML-діаграма взаємодії алгоритмів обробки даних з користувачем

Ключові можливості CLIP:

- семантичне узгодження візуальної інформації про наявні продукти з текстовими вподобаннями користувача;
- одночасне розуміння відповідностей між візуальними та текстовими концепціями;

- формування єдиного векторного простору для різнотипних даних;
- забезпечення більш точної та релевантної генерації рецептів порівняно з окремим використанням CNN-моделей та текстових кодувальників.

Даний набір алгоритмів дозволяє побудувати удосконалений алгоритм обробки візуальних кулінарних даних для подальшого використання в генерації персоналізованих рецептів, що є одним з основних принципів роботи даного модуля. UML-діаграма взаємодії алгоритмів обробки кулінарних даних з користувачем зображено на рисунку 2.1.

Висновки

Отже, запропонована структура алгоритму обробки кулінарних даних забезпечить підвищення точності створення релевантних рецептів завдяки використанню єдиного векторного представлення різнотипних вхідних даних (текстових запитів, зображень інгредієнтів, дієтичних параметрів), а також застосування мультимодальних моделей типу CLIP для семантичного узгодження візуальної та текстової інформації, що дозволяє підвищити персоналізацію та релевантність генерованих інструкцій приготування страв.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Ліу, Б. (2011). Семантичний пошук на основі онтологій. Видавництво Спрінгер.
2. Манінгер, Д., & Векслер, Д. (2019). Обробка природної мови для семантичного пошуку. Журнал досліджень в галузі штучного інтелекту, 35(2), 123-156.
3. Гупта, С., & Гупта, А. (2022). Машинне навчання для семантичного пошуку: сучасний стан та перспективи. Огляд інформатики та комунікацій, 18(4), 321-345.
4. Тан, М., & Ле, К. (2019). EfficientNet: Переосмислення масштабування моделей для згорткових нейронних мереж. Матеріали 36-ої міжнародної конференції з машинного навчання, 97, 6105-6114.

Паночішин Юрій Миколайович — доцент технічних наук, професор кафедри комп'ютерних наук Вінницький національний технічний університет, м. Вінниця, e-mail: y.panochyshyn@vntu.edu.ua

Коханівський Антон Павлович — студент групи 1КН-22м, факультет інтелектуальних інформаційних технологій та автоматизації Вінницький національний технічний університет, Вінниця, e-mail: balalauka62@gmail.com

Yurii Mykolaiovych Panochyshyn — associate professor of technical sciences, professor of the department of computer science, Vinnytsia National Technical University, Vinnytsia, e-mail: y.panochyshyn@vntu.edu.ua

Kokhanivskyi Anton Pavlovich — student of group 1kn-22m, faculty of intellectual information technologies and automation, Vinnytsia National Technical University, Vinnytsia, e-mail: balalauka62@gmail.com