ISSN 3083-7715 (online) *Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)*
ISSN 3083-7650 (0135-1710)(print) *Management Information System and Devices. 2026. No. 1 (188)*

УДК 004.75 DOI: https://doi.org/10.30837/0135-1710.2026.188.054

Петров К. Е., Божко О. Ю.

МЕТОД АВТОМАТИЧНОЇ РОЗМІТКИ ОЗНАК СТРУКТУРНОЇ СКЛАДНОСТІ ДОКУМЕНТІВ ІЗ ВИКОРИСТАННЯМ ВЕЛИКИХ МОВНИХ МОДЕЛЕЙ

Предметом дослідження *є методи автоматизованого аналізу й визначення ознак структурної складності слабкоструктурованих ділових документів із застосуванням сучасних мультимодальних великих мовних моделей.* ***Мета*** *– розробити й експериментально перевірити працездатність методу автоматичної розмітки ознак структурної складності документів, який забезпечує можливість прогнозування якості подальшої екстракції даних, а також побудувати математичну модель для оптимізації вибору конфігурації мовних моделей у просторі критеріїв "якість – вартість".* ***Завдання*** *передбачають формалізацію таксономії ознак, формування експериментального корпусу, розроблення уніфікованої схеми промптингу й розв'язання задачі багатокритеріальної оптимізації вибору моделей.* ***Методологічну основу дослідження*** *становлять методи системного аналізу, емпіричне профілювання великих мовних моделей, методи інженерії підказок (prompt engineering), методи математичної статистики (кореляційний аналіз, розрахунок метрик Precision, Recall, F1) для оцінювання якості класифікації, а також методи дискретної оптимізації для знаходження компромісних рішень.* ***Результати.*** *Основним результатом досліджень є розроблений оригінальний метод автоматичної розмітки ознак структурної складності документів, що дає змогу автоматично генерувати структурований профіль складності документа й обробляти його оригінальне візуальне PDF-подання. Також запропоновано зважену функцію якості, яка бере до уваги ступінь впливу кожної ознаки на помилки екстракції, та виділено Парето-оптимальні конфігурації, що уможливлюють мінімізацію витрат з огляду на задані вимоги до надійності. На підставі досягнутих результатів можна зробити певні* ***висновки.*** *Експериментально доведено, що економічно ефективні моделі в режимі з використанням прикладів забезпечують високу точність розмітки, яка конкурує з результатами значно вартісних моделей. Установлено, що застосування режимів із розширеним міркуванням для задачі бінарної класифікації є економічно недоцільним через диспропорційне зростання вартості без суттєвого приросту якості. Запропонований метод розв'язує проблему відсутності інструментів попереднього оцінювання складності в системах інтелектуального опрацювання документів. Він забезпечує прогнозованість і керованість процесів екстракції, даючи змогу реалізувати адаптивну маршрутизацію документів залежно від їх складності. Це створює підґрунтя для побудови ефективних промислових систем із збалансованими показниками точності та вартості експлуатації, усуваючи необхідність у трудомісткій ручній розмітці.*

Ключові слова: *слабкоструктурований документ; екстракція даних; промпт-інжиніринг; Парето-оптимальність; мультимодальні моделі.*

1. Вступ

Автоматизація опрацювання ділових документів є важливим напрямом розвитку сучасних інформаційних систем управління.

Попри досягнення у сфері інтелектуального аналізу текстів, процес екстракції структурованих даних із документів залишається нетривіальним завданням через різноманіття форматів, неузгодженість розмітки й наявність прихованих контекстуальних залежностей. Оцінювання структурної складності документа може слугувати основою для прогнозування якості подальшої екстракції даних і вибору відповідної стратегії їх оброблення. Традиційні підходи до визначення ознак складності передбачають експертний аналіз, що є трудомістким процесом, має елемент суб'єктивності й погано масштабується для великих документних корпусів.

Застосування великих мовних моделей (LLM) для розв'язання цієї проблеми є перспективним напрямом досліджень і створює передумови для автоматизованого розпізнавання таких ознак без попереднього навчання на спеціалізованих даних. Проте відсутність апробованих методів автоматичної розмітки ускладнює оцінювання надійності отриманих результатів і порівняння ефективності використання різних моделей.

Тому актуальним є розроблення методу автоматичної розмітки ознак структурної складності документів із використанням мультимодальних мовних моделей. Установлення кореляційних зв'язків між автоматично виявленими ознаками складності та

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

фактичними показниками якості екстракції даних (зокрема мірою F1) дасть змогу оцінити практичну придатність таких моделей для попереднього прогнозування складності опрацювання документів.

2. Аналіз літературних джерел і визначення проблеми дослідження

Вивчення структурної складності документів має тривалу історію, яка охоплює період від систем на основі правил до сучасних методів глибокого навчання. У ранніх роботах [1, 2] розглянуто методи, що ґрунтуються на застосуванні експертних правил і граматичних підходів для опису логічної структури документів. Однак суттєвими недоліками таких методів є чутливість до "шуму" й складність масштабування на нові типи макетів через необхідність ручного створення наборів правил. Подальші статистичні методи [3] й класичне машинне навчання [4] дали змогу частково автоматизувати розпізнавання структурних елементів, проте й надалі потребували великих обсягів навчальних даних і ручної розмітки.

З появою спеціалізованих моделей глибокого навчання, що використовуються для аналізу документів, дослідники досягли значного покращення точності аналізу складних макетів. Прикладом таких моделей є *TableNet* [5], призначена для виявлення таблиць, або серії моделей *LayoutLM* [6], де був реалізований механізм самоуваги з огляду на просторові відношення. Ці архітектури об'єднали текстові, візуальні та просторові ознаки, що дало змогу розуміти слабкоструктуровані документи, зокрема рахунки-фактури чи форми. Проте навіть найкращі моделі залишалися залежними від великих обсягів анотованих даних, а ручна розмітка таких корпусів залишається однією з основних проблем. Процес ручної розмітки вимагає значних ресурсів, а надійність розмітки залишається низькою для складних випадків.

У сучасних студіях [7–9] автори доводять, що структурна складність документів безпосередньо впливає на продуктивність систем інформації, зокрема на показник якості екстракції даних F1-міри (F1-*score*). Для документів із багаторівневою структурою, неоднозначними підписами або вкладеними таблицями спостерігається зменшення значення F1-міри до 50%, якщо порівнювати з простими форматами.

Це створює об'єктивну потребу в методах, які могли б попередньо визначати рівень складності документа ще до етапу екстракції даних.

Попри значні успіхи у сфері використання засобів штучного інтелекту для роботи з документами, сучасні підходи не розв'язують проблеми автоматичного оцінювання ознак структурної складності документів. Наявні системи аналізу макетів або потребують повної розмітки, або зосереджуються на виявленні елементів макета (заголовки, таблиці, списки), але не на концептуальних властивостях, що впливають на якість розпізнавання.

Отже, навіть найсучасніші методи не забезпечують універсального засобу кількісного прогнозування складності документа.

Останні дослідження у сфері синтезу великих мовних моделей відкрили нові можливості для когнітивного аналізу документів без попереднього навчання. Так, автори праць [10, 11] засвідчили, що LLM здатні узагальнювати контекст, відстежувати структуру та ідентифікувати логічні зв'язки в документі навіть у режимі без прикладів (*zero-shot mode*). Проте питання застосування LLM для автоматичної розмітки ознак складності документів залишається практично недослідженим. Особливо це стосується бізнес-документів, де різноманітність форматів і мовних структур створює суттєві труднощі для класичних моделей.

Крім того, у науковій літературі відсутні дослідження, що оцінюють узгодженість рішень між різними мовними моделями або ефективність ансамблевих (*ensemble)* підходів до розмітки документів. Тоді як мультиагентні архітектури [12, 13] демонструють переваги в задачах логічного міркування, їх застосування до завдань оцінювання складності ділових документів ще не було предметом системного аналізу.

Можемо підсумувати, що результати попередніх досліджень дають змогу сформулювати основну проблему: попри наявність розвинених методів аналізу макетів і візуально-текстового подання документів, не існує ефективних методів автоматичного визначення ознак структурної складності, які б поєднували переваги LLM, можливість роботи без ручної анотації та відтворюваність результатів. Це створює передумови для розроблення й порівняння ефективності нових методів, зокрема ансамблевих підходів із використанням розширеного мисленням (*reasoning*

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

models) як арбітра між рішеннями кількох моделей нижчого рівня.

3. Мета й завдання дослідження

Метою дослідження є розроблення та експериментальна перевірка методу автоматичної розмітки ознак структурної складності бізнес-документів у форматі PDF із використанням великих мовних моделей, а також побудова на цій основі формальної моделі вибору конфігурації моделей у просторі "якість – вартість".

Запропонований метод має забезпечити ймовірність отримання для кожного документа структурованого подання його складності, яке може використовуватися для прогнозування очікуваної якості екстракції слабкоструктурованих даних для адаптивного вибору моделей екстракції та для керованого компромісу між якістю і вартістю оброблення.

Для досягнення поставленої мети в роботі необхідно розв’язати такі завдання:

– уточнити й формалізувати таксономію ознак структурної складності бізнес-документів, виокремити підмножину ознак, найбільш релевантних для задачі екстракції слабкоструктурованих даних, й обґрунтувати критерії їх відбору для експериментальної перевірки методу;

– сформувати експериментальний корпус бізнес-документів у форматі PDF з ручною еталонною розміткою обраних ознак структурної складності та прив’язкою до результатів екстракції даних із попереднього дослідження, забезпечити цим достатнє різноманіття структур і типів складності;

– розробити метод автоматичної розмітки ознак структурної складності з використанням мультимодальних великих мовних моделей, визначити формат вхідної інформації (передача PDF-документів у вигляді бінарних вкладень), уніфіковану схему вихідних даних (рішення про наявність ознаки, текстове пояснення й числова оцінка впевненості), а також принципи побудови запитів і налаштування режимів роботи моделей;

– експериментально оцінити якість автоматичної розмітки ознак складності для кількох мовних моделей різних провайдерів у різних режимах роботи (без прикладів, з прикладами, з поглибленим внутрішнім міркуванням), порівняти результати з ручною розміткою й проаналізувати вплив окремих ознак на якість екстракції даних;

– побудувати зважену функцію якості для множини моделей, яка агрегує значення якості автоматичної розмітки за всіма розглянутими ознаками з огляду на сили їх зв’язку з результатами екстракції слабкоструктурованих даних, та на цій підставі сформулювати задачу оптимізації вибору підмножини моделей у просторі "якість – вартість";

– розв’язати поставлену задачу оптимізації на основі емпіричних даних про якість і вартість використання мовних моделей, виокремити Парето-оптимальні конфігурації моделей і описати правило спеціалізованого призначення моделей окремим ознакам структурної складності, яке забезпечує практично доцільний баланс між точністю розмітки, вартістю та прогнозованістю процесу екстракції даних.

4. Матеріали й методи дослідження

Для розв’язання задачі автоматичної розмітки ознак структурної складності документів запропоновано оригінальний метод на основі емпіричного профілювання ефективності великих мовних моделей.

Ключова гіпотеза методу полягає в тому, що різні моделі демонструють неоднакову ефективність на різних типах ознак складності, тому призначення кожній ознаці оптимальної моделі забезпечує вищу загальну точність за контрольованої вартості порівняно з універсальним підходом використання однієї моделі для всіх ознак.

Метод передбачає реалізацію трьох послідовних етапів.

1. Профілювання моделей на експериментальному корпусі з еталонною розміткою для визначення ефективності кожної конфігурації моделі на кожній ознаці складності та вимірювання обчислювальної вартості.

2. Оптимальне призначення моделей до ознак з огляду на точність, вартість і детермінізм результатів.

3. Застосування сформованої системи призначень до нових документів з паралельним обробленням різних ознак різними моделями.

Реалізація першого етапу полягає у формуванні експериментального корпусу з еталонною розміткою та в систематичному оцінюванні конфігурацій моделей за метриками якості та вартості. Такий корпус необхідний для профілювання ефективності моделей і визначення оптимальних конфігурацій.

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

В ідеальних умовах метод має бути перевірений на повному наборі ознак складності, виявлених у попередніх дослідженнях оброблення документів. Однак комплексна перевірка всіх можливих ознак пов'язана зі значними обчислювальними витратами й вимагає суттєвих ресурсів для профілювання моделей. Тому доцільним є відбір репрезентативної підмножини ознак, що забезпечує достатню валідацію методу за умови контрольованих витрат.

Для експериментальної перевірки методу ознаки обираються за такими критеріями:

1) емпірично підтверджений вплив на якість екстракції – ознака має демонструвати статистично значущу кореляцію з метриками якості екстракції даних;

2) покриття різних типів когнітивної складності – відібрані ознаки мають демонструвати різні аспекти складності документів (структурні, семантичні, контекстуальні), щоб забезпечити репрезентативність висновків;

3) достатнє подання в корпусі – кожна ознака має бути присутня в необхідній кількості документів для надійного оцінювання якості її автоматичної розмітки різними моделями;

4) обмеження дослідницьких ресурсів – кількість відібраних ознак має бути збалансована з доступними обчислювальними ресурсами й бюджетом на використання API комерційних моделей.

Розмір експериментальної вибірки документів з базового корпусу визначається на основі принципу стратифікованої випадкової вибірки зі збереженням розподілу ознак складності, що забезпечує репрезентативність результатів за умови скорочення обчислювальних витрат.

Конкретні параметри експерименту, як-от: кількість виокремлених ознак, розмір корпусу, властивості документів – подано в п'ятому розділі цієї статті.

Для профілювання ефективності автоматичної розмітки ознак складності використовуються великі мовні моделі, які задовольняють перелічені нижче вимоги. Моделі мають підтримувати безпосереднє оброблення PDF-файлів без попереднього перетворення в текстовий формат.

Ця вимога зумовлена тим, що значна частина ознак структурної складності (зокрема об'єднання комірок, візуальне розташування елементів, просторові залежності) закодована у візуальному поданні документа.

Попередня конвертація PDF у текст або HTML що ускладнює або унеможливлює коректну ідентифікацію окремих ознак складності. Мультимодальні моделі, що обробляють PDF нативно, зберігають повний контекст документа, зокрема візуальну структуру й макет, що є критичним для об'єктивного оцінювання їх здатності до розмітки ознак.

Моделі мають бути доступними через програмний інтерфейс (API) без необхідності локального розгортання інфраструктури. Ця вимога забезпечує практичну застосовність методу: результати можуть бути легко відтворені іншими дослідниками або імплементовані у виробничих системах без складного технічного налаштування. API-доступ також гарантує використання актуальних версій моделей і забезпечує масштабованість оброблення.

Окремий інтерес становить експериментальна перевірка моделей у режимах міркування (*thinking modes*), які забезпечують покрокове обґрунтування прийнятих рішень.

Попри те, що такі режими зазвичай не дають змоги встановлювати температуру на рівні нуля й, відповідно, не гарантують повної детермінованості результатів, вони ефективні для оцінювання впливу експліцитних процесів міркувань на точність розмітки складних ознак. Порівняння продуктивності стандартних режимів і режимів міркувань дає змогу виявити типи ознак, для яких покрокове міркування забезпечує суттєве покращення якості.

Якість автоматичної розмітки кожної бінарної ознаки складності оцінювалася способом порівняння результатів роботи моделей з еталонною ручною розміткою.

Для кожної ознаки $f_i\,(i=1,\ldots,m)$, де m – загальна кількість ознак складності, та кожної конфігурації моделі обчислювалися стандартні метрики бінарної класифікації.

Нехай для ознаки f_i та документа d відомі:

– $y_i \in \{0,1\}$ – еталонне значення (ручна розмітка);

– $y_i \in \{0,1\}$ – передбачене значення (результат роботи моделі).

На основі порівняння еталонних і передбачених значень для всіх документів корпусу визначаються базові величини:

– *TP* (*True Positives*) – кількість документів, де ознака правильно визначена як присутня $\left(y_i=1, y_i=1\right)$;

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

– *FP* (*False Positives*) – кількість документів, де ознака хибно визначена як присутня $\left(y_i = 0, y_i = 1\right)$;

– *TN* (*True Negatives*) – кількість документів, де ознака правильно визначена як відсутня $\left(y_i = 0, y_i = 0\right)$;

– *FN* (*False Negatives*) – кількість документів, де ознака хибно визначена як відсутня $\left(y_i = 1, y_i = 0\right)$.

На основі перелічених величин обчислюються чотири метрики якості.

1. *Precision* (точність) – частка правильних позитивних передбачень серед усіх позитивних передбачень:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}. \quad (1)$$

2. *Recall* (повнота) – частка правильно визначених позитивних випадків серед усіх еталонних позитивних випадків:

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}. \quad (2)$$

3. F1-міра – гармонічне середнє точності та повноти:

$$\text{F1}_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}. \quad (3)$$

4. *Accuracy* (точність класифікації) – частка правильних передбачень (як позитивних, так і негативних):

$$\text{Accuracy}_i = \frac{TP_i + TN_i}{N}. \quad (4)$$

Для інтегрального оцінювання якості роботи конфігурації моделі на всіх ознаках використовувався показник макроусередненої F1-міри. Відповідно до загальноприйнятої методології оцінювання мультикласової класифікації [14, 15] ця метрика розраховується як середнє арифметичне F1-мір для окремих класів (ознак), що дає змогу рівноправно брати до уваги внесок як поширених, так і рідкісних класів без зміщення, властивого зваженим методам.

$$\text{F1}_{macro} = (1/n)\sum_{i=1}^{n} \text{F1}_i, \; i = 1, \ldots, n. \quad (5)$$

Цей показник забезпечує збалансовану увагу до якості розмітки всіх ознак незалежно від їх подання в корпусі й використовується як основний критерій для порівняння ефективності різних конфігурацій моделей.

Другий етап методу полягає у формалізації та розв'язанні задачі оптимального вибору підмножини моделей для автоматичної розмітки ознак складності. На відміну від тривіального підходу, що передбачає використання однієї моделі для всіх ознак, запропонований метод дає змогу призначити кожній ознаці найбільш ефективну модель, формуючи спеціалізовану систему, що забезпечує оптимальне співвідношення між якістю розмітки й обчислювальними витратами.

На основі результатів профілювання маємо:

– множину моделей $M = \{m_1, m_2, \ldots m_k\}$, де k – кількість моделей;

– множину ознак складності $C = \{c_1, c_2, \ldots c_n\}$, де n – кількість ознак;

– значення F1_{ij} – F1-міра, досягнута моделлю m_i на ознаці c_j $(i = 1 \ldots k;\; j = 1 \ldots n)$;

– вектор вартості $V = (v_1, v_2, \ldots, v_k)$, де v_i – вартість використання моделі m_i для оброблення одного документа.

Критична особливість задачі полягає в тому, що модель m_i за одну операцію (один виклик API), яка має вартість (коштує) v_i, здійснює розмітку всіх n ознак одночасно, але з різною якістю F1_{ij} для кожної ознаки c_j. Це принципово відрізняє цю задачу від класичних задач призначення, де кожен ресурс використовується для одного об'єкта.

Запропонуємо бінарний вектор рішень $x = \{x_1, x_2, \ldots x_k\}$, де $x_i = 1$, якщо модель m_i застосовується, та $x_i = 0$ – в іншому випадку. Якість розмітки ознаки c_j за умови вибору підмножини моделей x визначається як максимум серед якостей усіх обраних моделей:

$$\text{F1}_j(x) = \max_{i \in \{1,\ldots,k\},\, x_i = 1} \text{F1}_{ij}. \quad (6)$$

Для порівняння різних конфігурацій моделей використовується скалярна функція якості $Q(S)$, яка агрегує значення F1 для всіх ознак структурної складності. У найпростішому випадку $Q(S)$ може задаватися як звичайне середнє арифметичне значення F1 за ознаками, що фактично передбачає рівну важливість усіх ознак.

Однак попередній аналіз кореляцій між ознаками структурної складності та загальною F1 екстракції даних показав, що окремі ознаки по-різному впливають на якість екстракції. Тому доцільно

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

використовувати зважений інтегральний показник, у якому ознаки із сильнішим зв'язком із F1 екстракції мають більший внесок у функцію якості.

Нехай C – множина ознак структурної складності. Для кожної ознаки $c \in C$ на попередньому етапі обчислюється коефіцієнт кореляції r_c між бінарною наявністю цієї ознаки та F1 якості екстракції даних (за результатами зовнішнього експерименту).

На основі модулів цих кореляцій задаються ваги:

$$w_c = \frac{|r_c|}{\sum_{g \in c} |r_g|}, \; c \in C, \tag{7}$$

які відтворюють відносну важливість кожної ознаки щодо її впливу на якість екстракції.

Позначаючи як $\mathrm{F1}_{S,c}$ якість розмітки ознаки f для множини моделей S (визначену як максимум F1 серед моделей із S), функцію якості для множини моделей $S \subseteq M$ визначатимемо таким чином:

$$Q(S) = \sum_{c \in C} w_c \cdot \mathrm{F1}_{S,c}. \tag{8}$$

Така конструкція відповідає практичному механізму роботи системи: для кожної ознаки $c \in C$ на етапі налаштування обирається модель із множини S, яка демонструє найвище значення $\mathrm{F1}_{m,c}$, і надалі саме ця модель використовується для автоматичної розмітки відповідної ознаки. Ваги w_c, отримані з таблиці кореляцій, забезпечують узгодженість цільового критерію $Q(S)$ з реальною значущістю ознак для задачі екстракції слабкоструктурованих даних: ознаки, що сильніше пов'язані зі зміною F1 екстракції, мають більший вплив на прийняття рішень щодо вибору конфігурації моделей.

Загальна вартість оброблення одного документа дорівнює сумі вартостей усіх обраних моделей:

$$\mathrm{Cost}(x) = \sum_{i=1}^{k} v_i \cdot x_i. \tag{9}$$

У наведених ознаках припускається, що для кожного документа розглядається фіксований перелік ознак структурної складності, а кожна модель з множини M під час одного запуску формує оцінки для всіх цих ознак одночасно. Це вказує на те, що вартість використання моделі не залежить від кількості ознак, які фактично будуть закріплені за цією моделлю: якщо бодай одна ознака призначена моделі m, документ обробляється нею повністю, і до функції вартості $\mathrm{Cost}(S)$ додається внесок c_m.

Отже, $\mathrm{Cost}(S)$ інтерпретується як сумарні витрати на оброблення одного документа всіма моделями, які містить множина S. Функція якості $Q(S)$, задана як середнє арифметичне значення максимумів $\max_{m \in S} \mathrm{F1}_{m,k}$ за ознаками, відтворює найкращий досяжний рівень F1 для кожної ознаки за умови, що для її розмітки застосовуються лише моделі з множини S. Максимум за моделями означає, що на етапі налаштування системи для кожної ознаки обирається відповідальна модель із S, яка демонструє найвище значення F1 для цієї ознаки, і надалі саме ця модель використовується для автоматичної розмітки відповідної ознаки в робочому режимі.

Сформульована задача є задачею багатокритеріальної оптимізації з двома конфліктуючими цілями: максимізацією якості $Q(x)$ та мінімізацією вартості $\mathrm{Cost}(x)$. Залежно від особливостей прикладної галузі та наявних обмежень ресурсів задачу вибору оптимальної підмножини моделей можна сформулювати в кількох варіантах. Розглянемо три основних.

Максимізація якості за умови обмеження на вартість. Застосовується, коли бюджет фіксований і необхідно отримати найкращу можливу якість розмітки. Математична модель має такий вигляд:

$$\max_x Q(x) \,|\, \mathrm{Cost}(x) \le V_{\{\max\}}, \; x_i \in \{0,1\}, \; i = 1, \ldots, k, \tag{10}$$

де $V_{\max}$ – максимально допустима вартість оброблення одного документа.

Мінімізація вартості за умови гарантії якості. Цей варіант доцільний, коли існують суворі вимоги до точності розмітки й необхідно досягти їх з мінімальними витратами:

$$\max_x \mathrm{Cost}(x) \,|\, \mathrm{F1}_j(x) \ge \tau_j, \; j = 1, \ldots, n, \; x_i \in \{0,1\}, \; i = 1, \ldots, k, \tag{11}$$

де τ_j – мінімально прийнятне значення F1-міри для j-ї ознаки.

Пошук Парето-оптимальних рішень. Загалом, коли важливо знайти компроміс, розглядається задача багатокритеріальної оптимізації. Множина рішень x^* називається Парето-оптимальною, якщо не існує іншого рішення x, що одночасно покращує обидва критерії, тобто $Q(x) \ge Q(x^*)$

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

та $\mathrm{Cost}(x) \le \mathrm{Cost}(x^*)$. Крім того, хоча б одна з нерівностей є строгою.

Задача належить до класу задач комбінаторної оптимізації з бінарними змінними й нелінійною цільовою функцією. За своєю структурою вона є узагальненням відомої задачі про покриття множини (*Set Cover Problem*) [16], ускладненої додатковим критерієм якості та "м'яким" характером покриття (через функцію максимуму).

Зважаючи, що отримана математична модель є типовою задачею дискретної оптимізації, її розв'язують відомими методами з використанням стандартних бібліотек (*OR-Tools, PuLP* тощо) без необхідності розроблення спеціалізованих алгоритмів.

Вихідним матеріалом для експерименту взято матеріал попереднього дослідження [11], в якому було сформовано корпус із 115 синтетичних інвойсів. Для них вручну виконано розмітку 11 бінарних ознак структурної складності документа та обчислено метрику якості екстракції даних F1 для кількох великих мовних моделей.

Аналіз кореляцій між наявністю кожної ознаки й значенням F1 показав, що лише частина критеріїв демонструє стійкий і статистично значущий зв'язок із якістю екстракції даних.

Для експериментального дослідження обрано чотири ознаки структурної складності, які є найбільш релевантними для задачі екстракції слабкоструктурованих даних з PDF-документів облікового призначення:

– наявність непідписаних даних у табличній частині (*has unlabeled data*);

– наявність неоднозначних або потенційно суперечливих полів (*has ambiguous fields*);

– наявність даних, що охоплюють кілька рядків чи стовпців таблиці (*has spanning data*);

– необхідність виконання арифметичних обчислень для відновлення відсутніх значень (*requires calculations*).

Вибір саме цієї підмножини ознак зумовлений такими міркуваннями: по-перше, ці ознаки демонструють найсильніший вплив на якість екстракції для всіх досліджуваних моделей; по-друге, вони покривають різні категорії складності (структурні, семантичні, контекстуальні); по-третє, їх автоматичне визначення становить задачу, що потребує когнітивного аналізу структури й змісту документа.

Узагальнені значення коефіцієнтів кореляції Пірсона між обраними ознаками й загальною F1-мірою якості екстракції даних для трьох моделей наведено в табл. 1.

Таблиця 1. *Коефіцієнти кореляції між ознаками структурної складності та F1-міри якості екстракції даних*

Ознака структурної складності документа	Значення F1-міри			Середнє |r|
	GPT-4.1-nano	*GPT-4o-mini*	*Gemini 2.5 Flash*	
has_ambiguous_fields	–0.298	–0.257	–0.469	0.341
has_spanning_data	–0.462	–0.431	–0.449	0.447
has_unlabeled_data	–0.477	–0.563	–0.597	0.546
requires_calculations	–0.331	–0.131	–0.277	0.246

Від'ємні значення r відповідають зменшенню F1-міри в присутності відповідної ознаки. Значущість коефіцієнтів оцінювалась за t-критерієм Стьюдента: для обсягу вибірки $N = 115$ та рівня значущості $\alpha = 0.05$ критичне значення кореляції становить $|r|_{crit} \approx 0.183$. Відповідно, усі коефіцієнти, що перевищують це значення за модулем, є статистично значущими.

Усі чотири ознаки мають помітний зворотний зв'язок з якістю екстракції: у присутності цих ознак значення F1, як правило, зменшується. Найбільше за модулем середнє значення $|r|$ спостерігається для ознаки непідписаних табличних даних і для ознаки даних, які охоплюють кілька рядків або стовпців, що свідчить про їх особливо несприятливий вплив на стабільність автоматизованого витягання реквізитів. Важливо, що частина обраних ознак принципово не може бути коректно визначена лише на основі лінійного текстового подання документа.

Ознаки, пов'язані з комірками, що охоплюють кілька рядків або стовпців, зі складеними комірками з кількома логічними полями чи непідписаними числовими даними, залежать від просторового розміщення елементів таблиці на сторінці. Для їх надійної розмітки модель має опрацьовувати структуроване або мультимодальне подання PDF-документа (табличну розмітку, координати, ієрархію блоків), а не лише "плоский" текст.

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

Отже, обраний набір ознак дає змогу не тільки оцінити вплив структурної складності на якість екстракції, але й протестувати здатність великих мовних моделей інтерпретувати повну візуально-структурну організацію документа.

Для експериментальної перевірки запропонованого методу автоматичної розмітки було сформовано підвибірку з 44 документів базового корпусу. До неї увійшли інвойси з еталонною (ручною) розміткою всіх 11 ознак структурної складності та відомими значеннями F1-міри якості екстракції для кожної моделі.

Підвибірка формувалася таким чином, щоб забезпечити достатню кількість прикладів позитивних і негативних значень для кожної з чотирьох обраних ознак та їх поєднань. Саме на цих 44 документах проводилося профілювання моделей та оцінювалась якість автоматичної розмітки ознак структурної складності.

5. Результати дослідження та їх обговорення

Розглянемо отримані результати автоматичної розмітки чотирьох ознак структурної складності для підвибірки з 44 PDF-документів.

Для кожного документа застосовувалися кілька конфігурацій великих мовних моделей трьох постачальників (*Anthropic Claude, OpenAI GPT-5, Google Gemini*) у різних режимах роботи: без прикладів у підказці, з використанням прикладів і за наявності підтримки так званих режимів міркування (*thinking mode*). Оцінювання виконувалося окремо для кожної з чотирьох ознак складності, а також за інтегральним показником – середнім арифметичним значенням F1-міри (*macro*-F1) за всіма ознаками для кожної конфігурації моделі.

У табл. 2 подано зведені значення F1-міри для кожної моделі, режиму застосування та ознаки, а також відповідні *macro*-F1, що дає змогу порівнювати якість розмітки як між різними постачальниками, так і між режимами для одних і тих самих моделей.

В експерименті розмітка ознак структурної складності виконувалася з використанням єдиного стандартизованого промпта. Усі документи передавалися до моделей як бінарні PDF-вкладення через API: окремо передавався PDF-файл, що підлягає розмітці, а також один або декілька PDF-файлів із прикладами. Для прикладів додатково надавалися окремі файли з еталонною розміткою ознак, що застосовувалися в промпті для демонстрації цільового формату відповіді та інтерпретації критеріїв складності.

Таблиця 2. *Оцінка якості автоматичної розмітки ознак структурної складності для різних моделей і режимів*

Режим	Модель	Значення F1-міри				
		macro	***has unlabeled data***	***has ambiguous fields***	***has spanning data***	***requires calculations***
ZeroShot	*claude-haiku-4.5*	0.456	0.065	0.708	0.764	0.286
	claude-sonnet-4.5	0.594	0.359	0.864	0.653	0.500
	gemini-2.5-flash	0.710	0.846	0.619	0.488	0.889
	gemini-3-pro	0.781	0.909	0.629	0.588	1.000
	gpt-5-mini	0.700	0.462	0.898	0.439	1.000
	gpt-5.1	0.618	0.500	0.780	0.692	0.500
	gpt-5.1codex	0.594	0.333	0.905	0.138	1.000
FewShot	*gemini 2.5 flash*	0.828	0.929	0.652	0.731	1.000
	gemini3 pro	0.771	0.966	0.414	0.703	1.000
	gpt-5-mini	0.660	0.400	0.846	0.562	0.833
	gpt-5.1	0.619	0.903	0.757	0.816	0.000
Thinking	*claude sonnet 4.5*	0.615	0.286	0.800	0.375	1.000
Thinking FewShot	*claude sonnet 4.5*	0.793	0.897	0.800	0.474	1.000
	gemini 3 pro	0.725	0.929	0.364	0.606	1.000

У всіх конфігураціях без використання режимів міркування (*thinking mode*) параметр температури (*temperature*) встановлювався на рівні 0, що забезпечувало максимально можливу детермінованість результатів для фіксованого набору вхідних файлів і тексту підказки. Це налаштування є принциповим, оскільки дає змогу інтерпретувати розбіжності в показниках якості як наслідок відмінностей між моделями та режимами, а не стохастичних варіацій генерації.

У тексті промпта явно задавалася вимога повертати результат розмітки у вигляді JSON-об’єкта

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

фіксованої структури. У відповіді формувався запис, що містив ідентифікатор документа (*document_id*) і словник критеріїв *criteria*, у якому для кожної ознаки структурної складності зберігався трійковий опис: *label, evidence* та *confidence*. Поле *label* могло набувати одного з трьох значень фіксованого словника (*present, absent, unsure*), тобто задавало тринарне рішення щодо наявності ознаки.

Для обчислення бінарних метрик якості (TP, FP, TN, FN, *Precision, Recall,* F1-міри) ці значення додатково відображалися у бінарну змінну: значення *present* тлумачилося як позитивний клас, а значення *absent* і *unsure* – як негативний.

Поле *confidence* містило числову оцінку впевненості моделі в діапазоні [0;1], а поле *evidence* – стисле текстове пояснення із посиланням на конкретні елементи PDF-документа, на підставі яких було прийнято відповідне рішення.

Такий формат відповіді, жорстко визначений промптом, є важливим як для інтерпретованості результатів (особливо в разі режимів міркування), так і для подальшої перевірки еталонної розмітки та ітеративного вдосконалення самого промпта.

Окрему групу конфігурацій становили так звані режими міркування (*thinking mode*) моделей із розширеним мисленням. Їх використання виявило низку практичних обмежень. По-перше, для режиму міркування *Anthropic Claude* мінімально доступне значення параметра *temperature* становило 1, а для *reasoning*-режиму *Gemini* – 0.2, що не дає змогу реалізувати повністю детермінований сценарій оброблення, аналогічний до режимів із $temperature = 0$. За таких налаштувань навіть для фіксованого набору вхідних PDF-файлів, прикладів і тексту підказки відповіді моделі залишаються стохастичними: повторні запуски можуть повертати різні варіанти значень *label, evidence* та *confidence* для тих самих документів.

По-друге, експериментальні результати продемонстрували, що використання режимів міркування (*thinking mode*) не забезпечує суттєвого виграшу в якості для задачі бінарної розмітки чотирьох ознак структурної складності. Значення F1-міри окремих ознак та *macro*-F1 для конфігурацій типу *Thinking* і *Thinking + FewShot* залишаються на рівні найкращих *FewShot*-конфігурацій або лише незначно їх перевищують, тоді як споживання токенів і, відповідно, вартість виконання запитів зростають у кілька разів.

У поєднанні з підвищеною недетермінованістю це робить *thinking*-режими менш привабливими для промислового застосування саме в розглянутій постановці задачі, де критичними є відтворюваність результатів і прогнозовані витрати, а структура рішення (формат JSON з *label, evidence* і *confidence*) уже забезпечує достатню інтерпретованість без необхідності додаткового багатоетапного міркування.

Водночас результати спостереження щодо роботи режимів міркування мають окрему методологічну цінність. Аналіз проміжних міркувань моделі, структури сформованих пояснень і властивих помилок дає змогу краще зрозуміти, як саме інтерпретуються формулювання критеріїв у промпті. Це розуміння може використовуватися для цілеспрямованого вдосконалення підказок: уточнення визначень ознак, добору більш репрезентативних прикладів, зміни інструкцій щодо оброблення пригранічних випадків.

Отже, незважаючи на те, що режими міркування в розглянутій задачі не є економічно доцільними для продуктового застосування, вони можуть розглядатися як корисний інструмент діагностики й налагодження промптів на етапі дослідження.

Для оцінювання економічної доцільності застосування різних конфігурацій моделей було проаналізовано споживання токенів і вартість виконання розмітки корпусу із 44 документів. Для кожної моделі окремо бралися до уваги кількість вхідних і вихідних токенів, тарифи постачальника за 1 млн токенів відповідного типу, а також підсумкова вартість оброблення всієї вибірки.

На підставі цих показників було обчислено повну вартість запуску кожної конфігурації моделі в межах проведеного експерименту. Зведені результати наведено в табл. 3, що дає змогу безпосередньо порівнювати моделі не лише за якістю розмітки ознак, а й за витратами, необхідними для досягнення такої якості.

У механізмі ціноутворення сучасних хмарних LLM принциповим є не тільки сам факт застосування тієї чи іншої моделі, а й спосіб її експлуатації: обсяг вхідних і вихідних токенів, наявність режимів міркування, використання кешування, а також під’єднання старших версій моделей із розширеним функціоналом.

У табл. 3 подано оцінку вартості для кожної конфігурації в припущенні застосування верхнього тарифного рівня для відповідного постачальника

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

та без огляду на механізми кешування. Це означає, що наведені значення відтворюють консервативну (максимальну) оцінку витрат у процесі оброблення 44 документів у межах експерименту, тоді як реальна вартість для малих обсягів може бути нижчою через дію тарифних градацій.

Таблиця 3. *Споживання токенів і вартість використання моделей*

Режим	Модель	Вхідні токени			Вихідні токени			Загальна вартість, USD
		кільк.	ціна, USD/1M	вартість, USD	кільк.	ціна, USD/1M	вартість, USD	
ZeroShot	*claude-haiku-4.5*	459680	0.100	0.046	25658	5.000	0.128	0.174
	claude-sonnet-4.5	459680	0.300	0.138	21579	15.000	0.324	0.462
	gemini-2.5-flash	130930	1.000	0.131	17194	2.500	0.043	0.174
	gemini-3-pro	301498	4.000	1.206	14380	18.000	0.259	1.465
	gpt-5-mini	340290	0.250	0.085	64232	2.000	0.128	0.214
	gpt-5.1	318090	1.250	0.398	16777	10.000	0.168	0.565
	gpt-5.1codex	318090	1.250	0.398	58229	10.000	0.582	0.980
FewShot	*gemini 2.5 flash*	432754	1.000	0.433	17512	2.500	0.044	0.477
	gemini3 pro	667336	4.000	2.669	16055	18.000	0.289	2.958
	gpt-5-mini	827852	1.250	1.035	60548	10.000	0.605	1.640
	gpt-5.1	740998	1.250	0.926	16554	10.000	0.166	1.092
Thinking	*claude sonnet 4.5*	461000	0.300	0.138	128791	15.000	1.932	2.070
Thinking	*claude sonnet 4.5*	852173	0.300	0.256	156324	15.000	2.345	2.601
FewShot	*gemini 3 pro*	657584	4.000	2.630	16325	18.000	0.294	2.924

Для моделей *Google* й *Anthropic* ціна за одиницю токена формально не залежить від того, використовується стандартний режим чи режим міркування. Однак доступ до такого режиму надається переважно в старших версіях моделей, що вже саме по собі підвищує базовий тариф, якщо порівнювати з "легкими" конфігураціями.

Крім того, у режимах міркування суттєво зростає кількість вихідних токенів: до звичайної відповіді моделі додаються токени, пов’язані з проміжними кроками міркування (ланцюжок думок, планування, розгортання структурованих пояснень тощо). У білінгу такі "токени міркування" розглядаються як вихідні, і для них застосовується тариф на вихідні токени, який у більшості комерційних планів істотно вищий, ніж для вхідних. Як наслідок, сумарні витрати для *reasoning*-конфігурацій можуть зростати в рази навіть за незмінної кількості оброблених документів.

Важливим елементом гнучкості є здатність обмежувати максимальну кількість токенів, що можуть бути витрачені під час оброблення одного запиту, зокрема й на етапі "обмірковування". Це дає змогу контролювати верхню межу вартості "міркувальних" запитів і запобігати неконтрольованому розширенню вихідної послідовності, однак водночас створює ризик передчасного обрізання ланцюжка міркувань і, зрештою, деградації якості відповіді.

Отже, налаштування лімітів для режимів міркування перетворюється на оптимізаційну задачу: необхідно досягти прийнятного компромісу між глибиною міркування, стабільністю якості та прогнозованістю витрат.

Додатковим чинником є тарифні градації великих постачальників: вартість 1 млн токенів зазвичай залежить від сумарного обсягу використання за певний період.

У табл. 3 для простоти порівняння взято до уваги максимальні (найгірші, на думку користувача) ставки. На практиці в сценаріях оброблення поодиноких документів або незначних партій реальна ціна за токен може бути нижчою, якщо обсяг застосування не досягає найвищих тарифних коридорів.

Водночас за умови масового оброблення значної кількості документів сумарні витрати різко зростають, і вплив вибору режиму (*ZeroShot, FewShot* або *Thinking*) і конкретної моделі стає критичним для економічної ефективності системи.

Окремо важливо наголосити на значенні механізму кешування токенів. У низці реалізацій LLM API передбачено можливість повторно використовувати попередньо зафіксовані фрагменти запитів (наприклад, системні інструкції, спільну частину промпта або незмінні приклади). До того ж вартість "кешованих" вхідних токенів значно нижча за повну ціну стандартного введення.

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

Для задачі розмітки ознак структурної складності це особливо актуально в сценаріях із застосуванням прикладів, де один і той самий набір еталонних прикладів багаторазово використовується для розмітки різних документів. Перенесення цих прикладів до кешу дає змогу суттєво зменшити сумарні витрати на вхідні токени. Водночас кешування практично не впливає на вартість вихідних токенів, а саме вихідні токени (зокрема й токени міркування) є найбільш вартісними.

Отже, механізм кешування ефективно знижує витрати в конфігураціях із значним обсягом повторюваного вхідного контексту, але майже не допомагає в разі, коли основна частина вартості зумовлена великою довжиною вихідних відповідей.

Загалом наведені фактори показують, що оцінювання економічної доцільності використання тієї чи іншої моделі не може обмежуватися лише порівнянням "номінальних" тарифів за 1 млн токенів. Необхідно зважати на розподіл навантаження між вхідними й вихідними токенами, наявність або відсутність режимів міркування, обсяги застосування, можливість кешування повторюваних фрагментів промпта, а також практичні обмеження на температурні налаштування й максимальну довжину відповіді.

Тільки з огляду на перелічені аспекти вартісний аналіз уможливлює коректне порівняння моделей між собою та обґрунтоване обрання конфігурації, що забезпечують найкращий баланс між якістю розмітки ознак структурної складності й витратами на обчислювальні ресурси.

Практичне використання системи автоматичної розмітки ознак структурної складності вимагає уваги не лише до досяжної якості, а й до вартості застосування великих мовних моделей.

Вище запропоновано зважену функцію якості $Q(S)$, яка агрегує F1-міри за всіма ознаками з використанням ваг w_c, пропорційних середнім за моделі абсолютним значенням коефіцієнтів кореляції $|r_c|$ між наявністю ознаки $c \in C$ і загальною $F1$ якістю екстракції даних.

Отже, ознаки, що демонструють сильніший зв'язок із деградацією або покращенням якості екстракції, мають більший внесок у цільовий показник $Q(S)$.

Для кожної моделі $m \in M$ та ознаки $c \in C$ за результатами експерименту було обчислено значення $F1_{m,c}$, що визначає якість автоматичної розмітки відповідної ознаки, а також емпіричну вартість повного прогону моделі c_m, встановлену на основі фактичного споживання вхідних і вихідних токенів. Для кожної підмножини моделей $S \subseteq M$ значення $F1_{S,c}$, зваженої функції якості $Q(S)$ та функції вартості $\mathrm{Cost}(S)$ обчислювалися за формулами (7)–(9), тобто як максимум F1 за моделями для кожної ознаки, зважена сума значень якості з використанням ваг w_c, отриманих із таблиці кореляцій, і сума вартостей c_m для всіх моделей, які містить множина S.

Така конструкція відповідає сценарію, у якому для кожної ознаки структурної складності на етапі налаштування системи обирається відповідна модель із множини S, що демонструє найвище значення $F1_{m,c}$. Саме цю модель надалі застосовано для автоматичної розмітки відповідної ознаки, тоді як вартість використання конфігурації визначається сумою вартостей усіх моделей, які хоча б для однієї ознаки беруть участь у розмітці: $\mathrm{Cost}(S) = \sum_{m \in S} c_m$. На завершальному етапі дослідження побудовано зважену функцію якості $Q(S)$ та функцію вартості $\mathrm{Cost}(S)$, які обчислюються за формулами (7)–(9), і сформульовано задачу вибору підмножини моделей у просторі "якість – вартість".

Для заданої множини моделей M необхідно знайти підмножину $S \subseteq M$, яка максимізує $Q(S)$ за умови, що загальна вартість $\mathrm{Cost}(S)$ не перевищує припустимого бюджетного порогу.

Після цього з множини всіх конфігурацій було виокремлено Парето-оптимальні (ефективні) набори моделей, для яких не існує іншого набору S', що має не гіршу якість за не більшу вартість, тобто $Q(S') \geq Q(S)$, $\mathrm{Cost}(S') \leq \mathrm{Cost}(S)$. Крім цього, хоча б одна з нерівностей є строгою.

Сукупність таких конфігурацій описує фронт Парето в просторі "якість – вартість" і задає практично релевантні компроміси між якістю автоматичної розмітки й витратами на використання моделей.

У табл. 4 наведено Парето-оптимальні конфігурації відповідно до зростання вартості. Для кожного набору моделей S подано сумарну

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

вартість $\mathrm{Cost}(S)$ (за результатами оброблення 44 документів) та зважену якість $Q(S)$, обчислену за описаною вище схемою (значення наведено з округленням до трьох знаків).

Отримані результати демонструють, що оптимально підібрані невеликі набори недорогих моделей допомагають забезпечити майже максимальну досяжну якість розмітки складності за значно нижчої вартості, ніж просте використання всіх доступних моделей, і без залучення вартісних режимів міркування.

Для кожної Парето-оптимальної конфігурації S, що наведені в табл. 4, відповідна модель для ознаки $c \in C$ визначається за правилом:

$$m^*(c; S) = \arg\max_{m \in S} F1_{m,c}. \qquad (12)$$

Таблиця 4. *Парето-оптимальні набори моделей у просторі "якість – вартість"*

№	Набір моделей, *S*	Вартість Cost(*S*), USD	Зважена якість, *Q*(*S*)
1	{ gemini-2.5-flash }	0.174	0.702
2	{ claude-haiku-4-5, gemini-2.5-flash }	0.348	0.800
3	{ gemini-2.5-flashF }	0.477	0.824
4	{ claude-haiku-4-5, gemini-2.5-flash, gpt-5-mini }	0.562	0.858
5	{ gemini-2.5-flashF, gpt-5-mini }	0.690	0.877
6	{ claude-haiku-4-5, gemini-2.5-flashF, gpt-5-mini }	0.864	0.887
7	{ gpt-5-mini, gpt-5.1F }	1.305	0.892
8	{ gemini-2.5-flashF, gpt-5-mini, gpt-5.1F }	1.782	0.901
9	{ gemini-2.5-flashF, gpt-5.1-codex, gpt-5.1F }	2.548	0.903
10	{ gemini-3-pro-previewF, gpt-5-mini, gpt-5.1F }	4.264	0.914
11	{ gemini-3-pro-previewF, gpt-5.1-codex, gpt-5.1F }	5.030	0.916

Іншими словами, для кожної ознаки обирається модель із множини S, яка демонструє найвище значення F1-міри для цієї ознаки, і саме ця модель використовується для її автоматичної розмітки.

Отже, табл. 4 дає не лише розуміння щодо оптимальних наборів моделей у просторі "вартість – якість", а й однозначно задає розподіл ознак між моделями для кожної обраної конфігурації.

6. Висновки й перспективи подальших досліджень

У межах проведеного дослідження розроблено метод автоматичної розмітки ознак структурної складності документів у форматі PDF з використанням великих мовних моделей. Метод може бути впроваджений для довільної множини ознак і довільного набору моделей і не обмежується певною кількістю критеріїв чи конкретними архітектурами. Ключовою ідеєю є те, що документ подається до мовної моделі у вигляді цілісного файлу, а на виході формується структурований опис його складності, який містить для кожної ознаки чітке рішення про її наявність, стисле текстове пояснення та числову оцінку впевненості в цьому рішенні.

Таке подання є універсальним і може безпосередньо використовуватися на подальших етапах оброблення документів. Запропонований метод робить процес екстракції слабкоструктурованих даних з PDF-документів не лише більш точним, а й прогнозованим і керованим. Оскільки для ознак структурної складності заздалегідь досліджено їх зв’язок із якістю екстракції, профіль складності конкретного документа може використовуватися як окрема вхідна властивість: за ним можна оцінити очікуваний рівень якості вилучення даних, визначити, чи достатньо застосувати недорогу модель, чи доцільно залучити більш потужну модель або передати документ на додаткову перевірку оператору.

Отже, розмітка ознак складності перетворюється на інструмент, за допомогою якого система екстракції набуває властивостей прогнозованості (очікуваний рівень якості відомий наперед) і керованості (користувач або розробник може явно обирати компроміс між якістю та витратами).

В експериментальній частині роботи метод застосовано до підмножини з чотирьох ознак структурної складності, для яких попередньо було встановлено найбільш виразний вплив на якість екстракції даних. Це лише один із можливих варіантів налаштування методу. Сам метод не прив’язаний ані до конкретної кількості ознак, ані до конкретних типів документів: за наявності

ISSN 3083-7715 (online) *Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)*
ISSN 3083-7650 (0135-1710)(print) *Management Information System and Devices. 2026. No. 1 (188)*

нових ознак або нових форматів документів достатньо повторити етап аналізу зв'язку між ознаками та якістю екстракції і оновити ваги в інтегральній функції якості. Автоматична розмітка ознак складності за допомогою великих мовних моделей у цьому разі дає змогу отримувати потрібні дані для будь-якого обсягу документів без різкого зростання витрат на ручну працю, тоді як повністю ручна розмітка такого самого обсягу була б значно вартісною та практично нереальною для підтримки в актуальному стані.

Експериментальні результати продемонстрували, що розроблений метод сприяє досягненню узгодженості автоматичної розмітки з ручною на рівні, достатньому для практичного використання в реальних конвеєрах оброблення документів. Запропонована зважена функція якості, яка бере до уваги різний вплив окремих ознак на підсумкову точність екстракції, у поєднанні з розв'язанням оптимізаційної задачі в просторі "якість – вартість" дала змогу побудувати Парето-оптимальні конфігурації мовних моделей.

Показано, що невеликі набори відносно недорогих моделей у режимах без прикладів і з незначною кількістю прикладів забезпечують майже максимально досяжну зважену якість автоматичної розмітки. Водночас конфігурації з режимами поглибленого внутрішнього міркування не продемонстрували переваги, яка б виправдовувала суттєве зростання витрат.

Це підтверджує, що автоматичну розмітку ознак складності доцільно будувати не на найскладніших і найдорожчих моделях, а на продуманій комбінації кількох більш доступних моделей, оптимально підібраних з огляду на співвідношення "якість – вартість".

Перспективи подальших досліджень пов'язані з розширенням таксономії ознак структурної складності, залученням ширшого кола типів документів та інтеграцією модуля автоматичної розмітки до промислових систем екстракції слабкоструктурованих даних. Профіль складності документа може слугувати основою для побудови моделей прогнозування якості екстракції, для адаптивного вибору конфігурації моделей залежно від складності конкретного документа й для моніторингу зміни потоку документів у часі.

Подальші дослідження можуть також бути присвячені розробленню багатокритеріальних моделей оптимізації, які одночасно беруть до уваги якість, вартість, затримку оброблення та стабільність результатів. Загалом це створює підґрунтя для побудови систем, у яких екстракція слабкоструктурованих даних з PDF-документів досягне необхідного рівня якості, а їх поведінка залишиться прозорою, прогнозованою та керованою навіть за умови постійних змін структури вхідних документів.

Конфлікт інтересів

Автори статті декларують, що не мають конфлікту інтересів, зокрема фінансового, особистого, авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій роботі.

Фінансування

Дослідження проводилося без фінансової підтримки.

Доступність даних

Рукопис не має пов'язаних матеріалів у сховищі даних.

Використання засобів штучного інтелекту

Автори підтверджують, що не застосовували технології штучного інтелекту для написання цієї роботи.

References

1. Document Parsing Unveiled: Techniques, Challenges, and Prospects for Structured Information Extraction. *arXiv*. 2024. URL: https://arxiv.org/html/2410.21169v1

2. Fisher, J. L. (1991), "Logical Structure Descriptions of Segmented Document Images", *Proceedings of International Conference on Document Analysis and Recognition (ICDAR)*, Saint-Malo, France, P. 302–310.

3. Vinay, V. et al. (2006), "Measuring the Complexity of a Collection of Documents", *Advances in Information Retrieval*, Vol. 3936, P. 107–118. DOI: https://doi.org/10.1007/11735106_11

4. Pembe, F. C., Güngör, T. (2015), "A Tree-Based Learning Approach for Document Structure Analysis and Its Application to Web Search", *Natural Language Engineering*, Vol. 21, No. 4, P. 569–605. DOI: https://doi.org/10.1017/S1351324914000023

5. Paliwal, S. et al. (2020), "TableNet: Deep Learning model for end-to-end Table detection and Tabular data extraction from Scanned Document Images", *arXiv*. DOI: https://doi.org/10.48550/arXiv.2001.01469

6. Huang, Y. et al. (2022), "LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking", *arXiv*. DOI: https://doi.org/10.48550/arXiv.2204.08387

7. Atagong, S. D. et al. (2025), "A review on knowledge and information extraction from PDF documents and storage approaches", *Frontiers in Artificial Intelligence*, Vol. 8, P. 1466092. DOI: https://doi.org/10.3389/frai.2025.1466092

8. Meuschke, N. et al. (2023), "A Benchmark of PDF Information Extraction Tools using a Multi-Task and Multi-Domain Evaluation Framework for Academic Documents", *arXiv*. DOI: https://doi.org/10.48550/arXiv.2303.09957

9. Bozhko, O. (2025), "Development of an iterative method for data extraction from unstructured documents based on the use of large language models", *Transactions of Kremenchuk Mykhailo Ostrohradskyi National University*, Iss. 1, P. 119–124. DOI: https://doi.org/10.32782/1995-0519.2025.1.15

10. Wang, W. et al. (2025), "Document Intelligence in the Era of Large Language Models: A Survey", *arXiv*. DOI: https://doi.org/10.48550/arXiv.2510.13366

11. Ding, Y. et al. (2025), "Deep Learning based Visually Rich Document Content Understanding: A Survey", *arXiv*. DOI: https://doi.org/10.48550/arXiv.2408.01287

12. Kupin, A. I., Kosei, M. P. (2024), "Overview of Multi-Agent System Architectures and Swarm Intelligence Algorithms", *Scientific notes of Taurida National V.I. Vernadsky University. Series: Technical Sciences*, Iss. 2, P. 98–104. DOI: https://doi.org/10.32782/2663-5941/2024.2/14

13. Петров, К. Е., Боков, І. П., Кобзев, І. В. (2025), "Розробка комбінованого методу аналізу емоційної забарвленості текстів", *АСУ та прилади автоматики,* Вип. 186, С. 5–16. DOI: https://doi.org/10.30837/0135-1710.2025.186.005

14. Jurafsky, D., Martin, J. H. (2024), "Speech and Language Processing (3rd ed. draft)". URL: https://web.stanford.edu/~jurafsky/slp3/ (accessed 20.08.2024).

15. Yamasaki, Chihiro, et al. (2025), "Function-based Labels for Complementary Recommendation: Definition, Annotation, and LLM-as-a-Judge", arXiv. DOI: https://doi.org/10.48550/ARXIV.2507.03945

16. Karp, R. M. (1972), "Reducibility Among Combinatorial Problems", *Complexity of Computer Computations*, P. 85–103.

Received (Надійшла) 30.11.2025
Accepted for publication (Прийнята до друку) 28.12.2025
Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Петров Костянтин Едуардович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, завідувач кафедри інформаційних управляючих систем; Харків, Україна;
Konstantin Petrov – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Head at the Information Control Systems Department; Kharkiv, Ukraine;
e-mail: kostiantyn.petrov@nure.ua
ORCID ID: https://orcid.org/0000-0003-1973-711X

Божко Олександр Юрійович – Харківський національний університет радіоелектроніки, аспірант кафедри інформаційних управляючих систем; Харків, Україна;
Oleksandr Bozhko – Kharkiv National University of Radio Electronics, Postgraduate Studentat the Information Control Systems Department; Kharkiv, Ukraine;
e-mail: oleksandr.bozhko@nure.ua
ORCID ID: https://orcid.org/0009-0004-6820-1228

METHOD OF AUTOMATIC LABELING FOR SIGNS OF STRUCTURAL COMPLEXITY IN DOCUMENTS USING LARGE LANGUAGE MODELS

The subject of the study *is methods of automated analysis and determination of structural complexity features in weakly structured business documents using modern multimodal large language models.* ***The purpose of the work*** *is to develop and experimentally*

ISSN 3083-7715 (online) *Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)*
ISSN 3083-7650 (0135-1710)(print) *Management Information System and Devices. 2026. No. 1 (188)*

test the performance of a method for automatic marking of signs of structural complexity in documents, which provides the ability to predict the quality of further data extraction, as well as to build a mathematical model for optimizing the selection of language model configurations in the "quality-cost" criteria space. ***The tasks*** *involve formalizing the taxonomy of features, forming an experimental corpus, developing a unified prompting scheme, and solving the problem of multi-criteria optimization of model selection.* ***The methodological basis*** *of the study consists of methods of system analysis, empirical profiling of large language models, prompt engineering methods, methods of mathematical statistics (correlation analysis, calculation of Precision, Recall, F1) for evaluating classification quality, as well as discrete optimization methods for finding compromise solutions.* ***Results****. The main result of the research is the development of an original method for automatic tagging of document structural complexity features, which allows automatically generating a structured document complexity profile and processing its original visual PDF representation. A weighted quality function is also proposed, which takes into account the degree of influence of each feature on extraction errors, and Pareto-optimal configurations are identified that enable cost minimization given the specified reliability requirements. Based on the results achieved, certain* ***conclusions*** *can be drawn. It has been experimentally proven that cost-effective models in example-based mode provide high marking accuracy that competes with the results of significantly more expensive models. It has been established that the use of extended reasoning modes for binary classification tasks is economically unfeasible due to a disproportionate increase in cost without a significant increase in quality. The proposed method solves the problem of the lack of tools for preliminary assessment of complexity in intelligent document processing systems. It ensures the predictability and controllability of extraction processes, enabling adaptive routing of documents depending on their complexity. This creates the basis for building effective industrial systems with balanced accuracy and operating costs, eliminating the need for labor-intensive manual marking.*

Keywords: *weakly structured document; data extraction; prompt engineering; Pareto optimality; multimodal models.*

Бібліографічні описи / Bibliographic descriptions

Петров К. Е., Божко О. Ю. Метод автоматичної розмітки ознак структурної складності документів із використанням великих мовних моделей. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 54–68. DOI: https://doi.org/10.30837/0135-1710.2026.188.054

Petrov, K., Bozhko, O. (2026), "Method of automatic labeling for signs of structural complexity in documents using large language models", *Management Information System and Devices*, No. 1 (188), P. 54–68. DOI: https://doi.org/10.30837/0135-1710.2026.188.054