DOI 10.31891/2307-5732-2024-339-4-37
УДК 004.62

БАСИСТЮК ОЛЕГ
Національний університет «Львівська політехніка»
https://orcid.org/0000-0003-0064-6584
e-mail: oleh.a.basystiuk@gmail.com
МЕЛЬНИКОВА НАТАЛІЯ
Національний університет «Львівська політехніка»
https://orcid.org/0000-0002-2114-3436
e-mail: nataliia.i.melnykova@lpnu.ua
ДУМИН ІРИНА
Національний університет «Львівська політехніка»
https://orcid.org/0000-0001-5569-2647
e-mail: iryna.b.shvorob@lpnu.ua
ДУМИН АНДРІЙ
Національний університет «Львівська політехніка»
https://orcid.org/0000-0003-2111-2899
e-mail: andrii.r.dumyn@lpnu.ua

# АНСАМБЛЕВИЙ ПІДХІД У МУЛЬТИМОДАЛЬНІЙ ОБРОБЦІ ДАНИХ НА ОСНОВІ GOOGLE API

*В роботі розглядаються сучасні тенденції, які використовуються у сфері обробки мультимодальних даних на основі ресурсів Google, основні напрямки розвитку сучасних методів комплексування даних. В процесі аналізу розглянути доцільність застосування методів побудови комунікаційних ліній із двома стратегіями ансамблювання.*

*Ключові слова: мультимодальні дані, перетворення мови в текст, розпізнавання мови, Sequence-to-Sequence, машинне навчання, штучний інтелект.*

BASYSTIUK OLEH, MELNYKOVA NATALIIA, DYMUN IRYNA, DYMUN ANDRII
Lviv Polytechnic National University

## ENSEMBLE APPROACH IN MULTIMODAL DATA PROCESSING BASED ON GOOGLE API

*This work examines current trends in multimodal data processing using Google resources and explores the main directions in developing modern data integration methods. The analysis considers the effectiveness of using communication lines with two ensembling strategies. Developing interfaces for multimodal data processing is a crucial step toward simplifying the analysis and interpretation of complex datasets. The research can be applied to develop speech-to-text models for various industries, enhancing speech translation tasks and boosting workers' time efficiency.*

*Context. Development of an interface for processing multimodal data using machine learning and an ensemble approach.*

*Objective. The propose a system architecture for the multimodal data processing interface that leverages modern ensemble approaches and machine learning techniques.*

*Methods. The proposed methodology is based on the integration of multiple models using communication lines to ensure a rapid and high-quality ensemble of data from different modalities. The architecture employs two main ensembling strategies: the A/B branching strategy and the sequential strategy.*

*Results. The proposed system architecture demonstrates a number of advantages:*

1. *Improved performance over traditional statistical machine translation systems that have been developed over the past two decades.*
2. *Independent of predefined language rules, algorithms are self-learning and frequently updated based on new data.*
3. *Efficient processing of multimodal data thanks to a flexible combination of ensemble strategies, which ensures efficient data integration and processing.*

*Conclusions. The proposed architecture, which combines branching and sequential communication lines, provides a robust framework for integrating various models, ensuring high-quality data analysis. This approach is promising for advancing multimodal data processing and offers significant potential for further research and development in the field.*

*Keywords: multimodal data, speech recognition, sequence-to-sequence, machine learning, artificial intelligence.*

### Постановка проблеми

Основна мета статті – розглянути і описати ключові етапи роботи інтелектуальної системи, зокрема, у цьому дослідженні розглядатимемо саме ансамблевий підхід, під час опрацювання мультимодальних даних на основі Google API. Обговорюються переваги та недоліки низки методологій, зокрема, заснованих на правилах, статистичних даних і нейронних мережах. Визначено найбільш підходящу програмну техніку та організаційну структуру для розробки рішень для оцінювання мультимодальних даних. Додатково, було розглянуто і запропоновано архітектурне вирішення для обробки мультимодальних даних, а також опрацьовано метод побудови комунікаційних ліній із двома стратегіями ансамблювання.

Наступним кроком у розвитку цього дослідження може стати реалізація пропонованої моделі для отримання фактичних результатів на різних наборах мультимодальних даних.

Для досягнення цієї мети були визначені наступні основні завдання дослідження:

- Проаналізувати існуючі методи і техніки для обробки мультимодальних даних;
- Представити мультимодальний інтерфейс обробки даних на основі рішень Google Cloud;
- Огляд використання техніки комунікаційних ліній на основі різних методів ансамлювання.

**Аналіз останніх джерел**

У попередніх дослідженнях розглядалися кілька ключових аспектів обробки мультимодальних даних, основним акцентом під час обробки даних зверталося на ефективність, надійність та безпечність. Це включає в себе пошук ефективних методів комплексування одномодальних та розділення мультимодальних даних, пошуку в системах шляхів для оптимізації, зокрема за допомого підбору коефіцієнтів та гіперпараметрів, що є важливим при використанні різних моделей машинного навчання. Останні джерела підтверджують ефективність сучасних методів машинного навчання для мультимодальної обробки даних [1-2].

Використання нейронних мереж, великих наборів даних та потужних обчислювальних ресурсів дозволяє значно покращити точність та швидкість аналізу. Крім того, стратегії ансамблювання, такі як комунікаційні лінії, забезпечують ефективне поєднання даних з різних модальностей, що підвищує якість кінцевих результатів. Важливість розглянути і запропонованих архітектури вирішень та ансамблевих методів обробки мультимодальних даних, є перспективними для подальшого розвитку у сфері обробки даних. Досліджуючи способи і методи обробки мультимодальних даних, зокрема української мови, дозволить створювати комплексні системи для аналізу широкого спектру проблеми пов'язаних з опрацювання природніх мов. Додатково, дослідження даної тематики відіграє важливу роль у процесі узгодження та підтримки інтеграції різних наборів даних (модальностей) між собою у системах опрацювання та інтерпретації даних.[3-4]

**Виклад основного матеріалу**

Розробка інтерфейсу для обробки мультимодальних даних є важливим кроком на шляху до полегшення аналізу та інтерпретації складних мультимодальних даних. Він дозволяє дослідникам і практикам об'єднувати дані з різних джерел і модальностей, щоб отримати глибше розуміння складних явищ. У наступних розділах ми опишемо підхід до майбутньої системної архітектури, який складається і представлений на рис. 1.

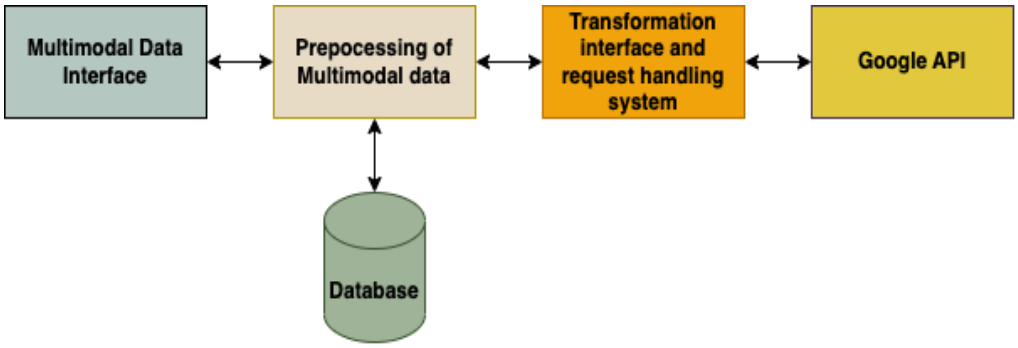


**Рис. 1. Мультимодальний інтерфейс обробки на основі Google API**

Основні переваги методу [5-7]:

- Система не залежить від знання будь-яких законів мови. Ці вказівки встановлюються самим алгоритмом і часто оновлюються;
- Запропонована методологія обмежена розміром набору навчальних даних і обсягом ресурсів обробки, які можуть бути спрямовані на переклад. Дослідники машинного навчання створили цей метод лише кілька років тому, але такі системи вже працюють краще, ніж статистичні системи машинного перекладу, які розвивалися протягом останніх 20 років.

Оскільки при опрацюванні та аналізі мультимодальних даних передбачається робота з кількома різними моделями, доцільним є використання техніки комунікаційних ліній, для забезпечення швидкого та якісного ансамблювання даних різних модальносте [8-10].

При побудові комунікаційних ліній застосовують дві стратегії ансамблювання:

1. Стратегія розгалуження A/B: стратегія, для якої одна частина вхідних даних надходить до однієї моделі, а інша – до другої моделі (якщо використовуються дві моделі).
2. Стратегія послідовності: вхідні дані проходять всі моделі одна за одною, з можливістю зупинки роботи на будь-якому кроці, якщо умови якості не були виконані.

Для розроблюваної системи доцільним буде застосування комбінації стратегій розгалуження A/B та послідовності. Запропонована схема роботи комунікаційних ліній зображена на рисунку 2.

Зі схеми видно, що весь процес ансамблювання даних розділено на дві комунікаційні лінії:

1. комунікаційна лінія A/B, що складається з:
   - компоненти для поділу даних в залежності від їх модальності між двома моделями Speech-to-Text та Natural Language Processing;
   - компонента для опрацювання даних за допомогою модуля Natural Language Processing;
   - комунікаційної лінія Sequence, що опрацьовує аудіодані;
   - компоненти постопрацювання вихідних даних, для їх зведення до одного вигляду;
2. комунікаційна лінія Sequence, що складається з:

- компонента для опрацювання даних за допомогою модуля Speech-to-Text;
- компоненти для фільтрування даних – частина отриманих результатів може бути готовою для зведення даних до одного вигляду, а частина даних потребує більш ретельного опрацювання;
- компонента для опрацювання даних за допомогою модуля AutoMl;
- компоненти постопрацювання вихідних даних, для їх зведення до одного вигляду.

Даний підхід дозволяє завантажити кілька моделей і зіставити їх з ключами за допомогою обробника моделі. Відображення моделей на ключі дає можливість використовувати різні моделі в одній комунікаційній лінії та зіставляти зведені дані для подальшої передачі в сховище даних.

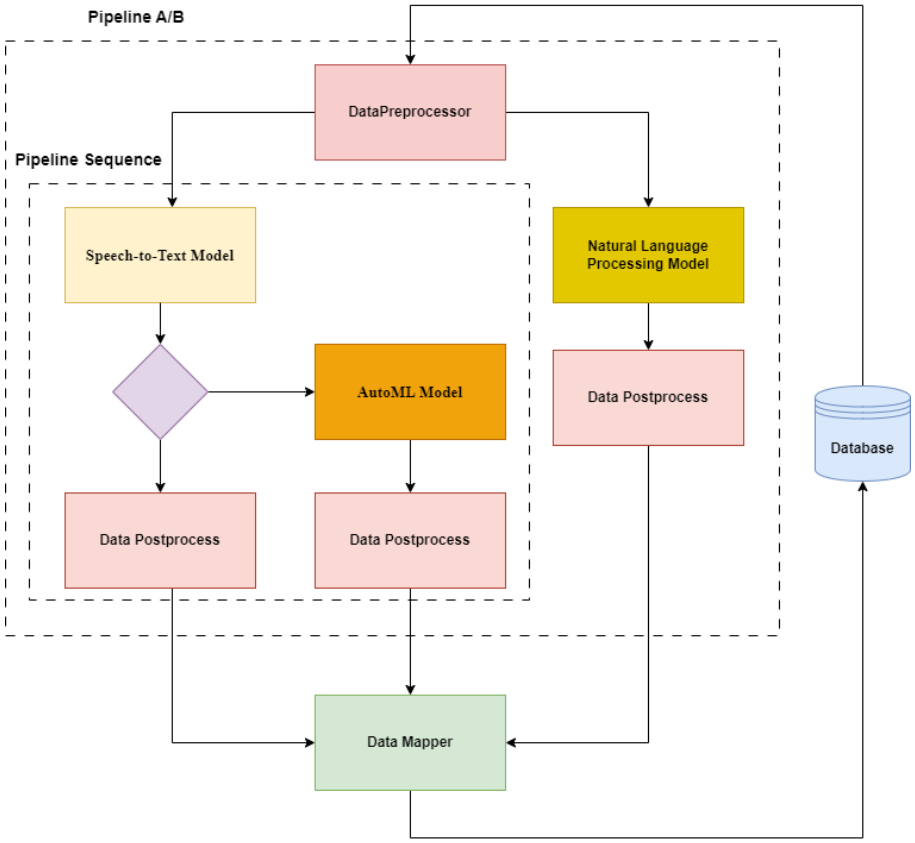


**Рис. 2. Запропонована схема роботи комунікаційних ліній**

## Висновки

Розробка інтерфейсу для обробки мультимодальних даних є важливим кроком у напрямку полегшення аналізу та інтерпретації даних, зокрема використовуючи інструменти на основі Google Cloud. Цей підхід дозволяє агрегувати дані з різних джерел і модальностей, що сприяє отриманню ґрунтовнішого розуміння складних явищ. Для ефективного аналізу мультимодальних даних пропонується використання техніки комунікаційних ліній, які забезпечують швидке та якісне ансамблювання даних різних модальностей. Запропонована комбінація стратегій розгалуження A/B та послідовності дозволяє ефективно опрацьовувати дані за допомогою різних моделей, таких як Speech-to-Text та Natural Language Processing. Запропонована методологія, дозволяє завантажувати кілька моделей, зіставляти їх з ключами та використовувати різні моделі в одній комунікаційній лінії, показала значний прогрес у порівнянні зі статистичними системами машинного перекладу. Важливою перевагою є те, що система не залежить від знання будь-яких законів мови, оскільки ці вказівки встановлюються самим алгоритмом і часто оновлюються. Це забезпечує ефективне зведення даних для подальшої передачі в сховище даних, що значно підвищує якість аналізу та інтерпретації мультимодальних даних.

## References

1. Jiang, Y., Irvin, J., Wang, J. H., Chaudhry, M. A., Chen, J. H., & Ng, A. Y. (2024). Many-Shot In-Context Learning in Multimodal Foundation Models. arXiv preprint arXiv:2405.09798.

2. M. Havryliuk, I. Dumyn, O. Vovk. (2023). Extraction of Structural Elements of the Text Using Pragmatic Features for the Nomenclature of Cases Verification. In: Hu, Z., Wang, Y., He, M. (eds) Advances in Intelligent Systems, Computer Science and Digital Economics IV. CSDEIS 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 158. Springer, Cham. DOI: https://doi.org/10.1007/978-3-031-24475-9_57.

3. C. Wang, N. Shakhovska, A. Sachenko, M, Komar. (2020). A new approach for missing data imputation in big data interface. Information Technology and Control, 49(4), pp. 541-555.

4. O. Basystiuk, N. Melnykova and Z. Rybchak, "Multimodal Learning Analytics: An Overview of the Data Collection Methodology," 2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT), Lviv, Ukraine, 2023, pp. 1-4, doi: https://doi.org/10.1109/CSIT61576.2023.10324177.

5. N. Melnykova, U. Marikutsa, Y. Kryvenchuk. (2018). The new approaches of heterogeneous data consolidation. In 2018 IEEE 13th international scientific and technical conference on computer sciences and information technologies (CSIT), Vol. 1, pp. 408-411.

6. K. Shakhovska, I. Dumyn, N. Kryvinska, M. K. Kagita, "An Approach for a Next-Word Prediction for Ukrainian Language", Wireless Communications and Mobile Computing, vol. 2021, 2021. DOI: https://doi.org/10.1155/2021/5886119

7. I. Zheliznyak, Z. Rybchak, I. Zavuschak, Analysis of clustering algorithms, 2017. Advances in Intelligent Systems and Computing, 2017, pp. 305 314.

8. Fedushko, S., Molodetska, K., & Syerov, Y. (2023). Analytical method to improve the decision-making criteria approach in managing digital social channels. Heliyon, 9(6).

9. O. Basystiuk, N. Melnykova. “Мультимодальне розпізнавання мовлення на основі звукових і текстових даних” / O. Basystiuk, N. Melnykova // Вісник Хмельницького національного університету. Технічні науки. – 2022. – № 5 (313). – С. 22–25.

10. N. Shakhovska, V. Bilynska, O. Syvokon, O. Shamura-tov, et. al.: "The Developing of the System for Autimatic Audio to Text Conversion", IT&AS'2021: Symposium on Information Technologies and Applied Sciences, March 5-6, 2021, Bratislava, Slovak Republic.