ISNN 3030-3907

2026/3
VOLUME 5

DEVELOPMENT OF SCIENCE

ILMIY JURNAL



Development of science

Ilmiy jurnal

2026/3 VOLUME 5

ISSN 3030 -3907

Ilmiy jurnal OAK rayosatining 2024 yil 27 sentabrdagi 361-son qaroriga asosan 02.00.00 kimyo fanlari hamda 2025 yil 12 fevral 367- son qaroriga asosan 05.00.00 texnika fanlari bo‘yicha falsafa doktori (PhD) va fan doktori (DSc) ilmiy darajasiga talabgorlarning dissertatsiya ishlari yuzasidan dissertatsiyalari asosiy ilmiy natijalarini chop etish tavsiya etilgan milliy ilmiy nashrlar ro‘yxatiga kiritilgan.

BOSH MUHARRIR:
Temirov Alisher Hoshim o‘g‘li – texnika fanlari falsafa doktori (PhD),Dotsent.
TAHRIRIYAT KENGASH RAISI:
Fozilov Sadriddin Fayzullayevich – texnika fanlari doktori, professor
TAHRIRIYAT KENGASH A’ZOLARI

TEXNIKA FANLARI/ TECHNICAL SCIENCES/ ТЕХНИЧЕСКИЕ НАУКИ

Bozorov G‘ayrat Rashidovich
Buxoro Davlat Texnika Universiteti, professor
Abdurahmonov Olim Rustamovich
Buxoro Davlat Texnika Universiteti, professor
Adizov Bobur Zamirovich
O‘zbekiston Respublikasi Fanlar akademiyasi Umumiy va noorganik kimyo instituti, professor
Hayitov Ruslan Rustamjonovich
Buxoro Davlat Texnika Universiteti, professor
Jumayev Qayum Karimovich
Buxoro Davlat Texnika Universiteti, professor.
Berdiyev Dorob Murodovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor
Maxmudov Rafiq Amonovich
Buxoro Davlat Texnika Universiteti, professor.
Tilloyev Lochin Ismatilloyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Raxmonov Qahramon Sanakulovich
Buxoro Davlat Texnika Universiteti, dotsent.
Barakayev Nusratullo Rajabovich
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti, professor.
Jo‘rayev Xayrullo Fayziyevich
Buxoro Davlat Texnika Universiteti, professor.
Panoyev Erali Rajabboyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Ataullayev Sherzod Nabiullyevich
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Malik Negmurodovich
Buxoro Davlat Texnika Universiteti, dotsent.
Axmedova Ozoda Bahronovna
Islom Karimov nomidagi Toshkent Davlat Texnika Universiteti.
Safarov Bahri Jumayevich
Buxoro Davlat Texnika Universiteti, dotsent.
Xaydarov Axtam Amonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Safarov Jasur Alijon o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Асадова Раъно Дилмуратовна
Тошкент Давлат техника университети доценти

KIMYO FANLARI/ CHEMICAL SCIENCES/ ХИМИЧЕСКИЕ НАУКИ

Axmedov Voxid Nizomovich
Buxoro Davlat Texnika Universiteti, professor
Umarov Baqo Bafoyevich
Buxoro Davlat Universiteti, professor
Kodirov Abduaxad Abduraximovich
Qarshi davlat Universiteti, professor
Qurbonov Mingniqul Jumagulovich
Qarshi davlat Universiteti, dotsent
Turemuratov Sharibay Naurizbaevich
O‘zbekiston Respublikasi Fanlar akademiyasi Qoraqalpog‘iston bo‘limi Bosh ilmiy kotibi, professor
Tursunov Kaxor Shonazarovich
Qarshi davlat Universiteti, professor
Do‘stov Hamro
Buxoro Davlat Texnika Universiteti, professor
Panoyev Nodir Shavkatovich
Buxoro Davlat Texnika Universiteti, dotsent.
Olimov Bobur Bahodir o‘g‘li
Buxoro Davlat Texnika Universiteti, dotsent.
Kasimov Sherzod Abduazimovich
Termiz Davlat Universiteti, professor.
Aliqulov Rustam Valiyevich
Termiz Davlat Universiteti, professor
Tropov Xamza Tursunovich
Samarqand Davlat Universiteti, professor

IQTISODIYOT FANLARI/ ECONOMICS/ ЭКОНОМИЧЕСКИЕ НАУКИ

Nizamov Aslitdin Badreddinovich
Buxoro Davlat Texnika Universiteti, professor.
Boboyev Akmal Choriyevich
Buxoro Davlat Texnika Universiteti, professor.
Ergashxojayeva Shaxnoza Djasurovna
Toshkent davlat iqtisodiyot universiteti, professor
Saidaxmedova Nodira Ilxomovna
Toshkent davlat iqtisodiyot universiteti, professor
Muxsinov Bekzod Toxirovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shakarov Qulmat Ashirovich
Renessans ta'lim universiteti,dotsent.
Jumaeva Zulfiya Qayumovna
Buxoro Davlat Texnika Universiteti, dotsent.
Norova Salomat Yusupovna
Buxoro Davlat Texnika Universiteti, dotsent.
Avezova Shaxnoza Maxmudjonovna
Buxoro Davlat Texnika Universiteti, dotsent.
Jalilov Jamshid G‘anijonovich
Toshkent davlat iqtisodiyot universiteti, dotsent
Xasanova Gulruh Djumanazarovna
Buxoro Davlat Texnika Universiteti, dotsent.

PEDAGOGIKA FANLARI/ PEDAGOGICAL SCIENCES / ПЕДАГОГИЧЕСКИЕ НАУКИ

Siddiqova Sadoqat G‘afforovna
Buxoro Davlat Texnika Universiteti, dotsent.
Zoyirov Erkin Xalilovich
Buxoro Davlat Texnika Universiteti, DSc, professor
Gaffarov Feruz Hasanovich
Puchon Universiteti, *professor*
Ruziyev Davron Yuldashevich
Buxoro Davlat Universiteti, professor
Muradova Firuza Rashidovna
Buxoro Davlat Texnika Universiteti, professor.
Yunusova Gulandon Samiyevna
Buxoro Davlat Texnika Universiteti, professor.
To‘xtayeva Zebo Sharifovna
Buxoro Davlat Texnika Universiteti, professor.
Navro‘zova Gulchixra Nigmatovna
Buxoro Davlat Texnika Universiteti, dotsent.
Uzoqova Laylo Polvonovna
Buxoro muhandislik – texnologiya instituti, dotsent.
To‘raqulova Marjona Qiyom qizi
Buxoro Davlat Texnika Universiteti, dotsent.
Muradova Zarina Rashidovna
Buxoro Davlat Texnika Universiteti, dotsent.
Murodov Sanjar Aslonovich
Buxoro Davlat Texnika Universiteti, dotsent.
Shodiyev Jahongir Jo‘raqulovich
Buxoro Davlat Texnika Universiteti, dotsent
Gaffarov Laziz Hasanovich
Buxoro muhandislik – texnologiya instituti, dotsent.

Tahririyat a’zolarning to‘liq ro‘yxati jurnalning www.devos.uz saytida keltirilgan.

Tahririyat hay’ati a’zolarining tarkibini o‘zgartirish hamda qo‘shimcha kiritish yoki tahririyat a’zoligidan chiqarish bosh muharrir tomonidan amalga oshiriladi.

ISSN 3030-3907

Development of science ilmiy jurnali 2024 yil 15 aprel kuni №256445 sonli guvohnoma bilan Davlat ro‘yxatidan o‘tgan.

To‘plam elektron ko‘rinishda (pdf formatida) mualliflarga taqdim etiladi. To‘plamga kiritiladigan maqolalarning mazmuni, ulardagi statistik ma’lumotlar va me’yoriy hujjatlarning to‘g‘riligiga hamda har qanday fikr - mulohazalarga mualliflarning o‘zlari mas’ul hisoblanadilar. Talablarga javob bermaydigan maqolalar to‘plamga kiritilmaydi. Tashkiliy qo‘mita maqola matnini qisqartirish, qisman tuzatish kiritish, seksiyalarga joylashtirish huquqiga ega.

ISSN 3030-3907

ADAPTIVE DISTRIBUTED NLP PROCESSING (ADNLP) ALGORITHM FOR LARGE-SCALE UZBEK LANGUAGES TEXTS

Hushnudbek Adinaev
Department of CE, Urgench State University, Urgench, Uzbekistan
0000-0002-7824-1854
Shahzodbek Ganijonov
10th grade, school 15, Urgench, Uzbekistan
0009-0003-7683-6998

Abstract. This paper proposes an Adaptive Distributed NLP Processing (ADNLP) algorithm for efficient punctuation restoration in large-scale Uzbek text corpora. Traditional BERT-based models require significant computational resources when processing large datasets. The proposed approach integrates Apache Spark–based distributed computing with dynamic partitioning and GPU-aware scheduling mechanisms. Experimental results show that the ADNLP algorithm achieves several times faster performance compared to the conventional sequential processing approach.

Keywords: Uzbek NLP, punctuation restoration, BERT, BiLSTM, Apache Spark, distributed computing.

Аннотация. В данной статье предлагается алгоритм адаптивной распределённой обработки естественного языка (ADNLP) для эффективного восстановления пунктуации в крупномасштабных узбекских текстовых корпусах. Традиционные модели на основе BERT требуют значительных вычислительных ресурсов при обработке больших объёмов данных. Предлагаемый подход объединяет распределённые вычисления на базе Apache Spark с механизмами динамического разбиения и планирования, ориентированного на использование GPU. Экспериментальные результаты показывают, что алгоритм ADNLP обеспечивает в несколько раз более высокую производительность по сравнению с традиционным последовательным подходом обработки.

Ключевые слова: узбекский NLP, восстановление пунктуации, BERT, BiLSTM, Apache Spark, распределённые вычисления.

Annotatsiya. Ushbu maqolada katta hajmdagi o‘zbek matn korpuslarida punktuatsiyani samarali tiklash uchun Adaptiv taqsimlangan NLP ishlov berish (ADNLP) algoritmi taklif etiladi. An’anaviy BERT asosidagi modellar katta hajmdagi ma’lumotlarni qayta ishlashda sezilarli hisoblash resurslarini talab qiladi. Taklif etilayotgan yondashuv Apache Spark asosidagi taqsimlangan hisoblashni dinamik bo‘lish (partitioning) va GPUga moslashtirilgan rejalashtirish mexanizmlari bilan birlashtiradi. Tajriba natijalari shuni ko‘rsatadiki, ADNLP

ISSN 3030-3907

algoritmi an’anaviy ketma-ket ishlov berish usuliga nisbatan bir necha baravar tezroq ishlash samaradorligini ta’minlaydi.

Kalit so‘zlar: o‘zbek NLP, punktuatsiyani tiklash, BERT, BiLSTM, Apache Spark, taqsimlangan hisoblash.

I. INTRODUCTION.

Natural Language Processing (NLP) has rapidly advanced in recent years with the emergence of deep learning models such as BERT. However, processing large-scale text corpora remains computationally expensive. This challenge is particularly significant for low-resource languages such as Uzbek, where efficient processing systems are still insufficiently developed. Punctuation restoration plays an important role in improving text readability and enabling downstream NLP tasks. Although transformer-based models provide high accuracy, their computational complexity limits large-scale deployment. To address these challenges, this paper proposes an Adaptive Distributed NLP Processing (ADNLP) algorithm that combines distributed computing with adaptive scheduling techniques. The proposed approach efficiently processes large-scale text corpora using Apache Spark while maintaining model performance.

II. RELETED WORK

1. Foundations of Large-Scale Data and Distributed Computing

The rapid growth of digital data necessitates a transition from traditional processing methods to distributed architectures in computational linguistics. Wigan and Clarke define Big Data as data volumes that cannot be effectively managed by conventional systems [1]. Experiments conducted by Omar and Jumaa demonstrate that the Apache Spark framework performs several times faster than Apache Hadoop due to its in-memory computing capabilities [2]. However, studies by Andersson indicate that in smaller clusters and simple Named Entity Recognition (NER) tasks, MapReduce can sometimes achieve performance comparable to Spark [3].

2. Uzbek Morphology and Approaches in Computational Linguistics

Uzbek, as an agglutinative language, differs fundamentally from the analytic structure of English, where grammatical meanings are encoded through a regular chain of affixes [4]. The UzMorphAnalyser model developed by Salaev analyzes agglutination using rule-based methods and achieves 91% accuracy [5]. Abdurakhmonova et al. applied Conditional Random Fields (CRF) and neural network models for morpheme segmentation and proposed a newly annotated dataset for the Uzbek language [6].

3. Modern Distributed Frameworks: Ray and Dask

ISSN 3030-3907

According to Walner's analysis, while Apache Spark is strong in data parallelism, the Ray framework offers advantages in logical parallelism and heterogeneous CPU/GPU resource management. This feature is important for optimizing neural network training and serving in the ADNLP system [7].

4. Adaptive Systems and Algorithmic Analysis

Adaptive systems are characterized by their ability to autonomously respond to environmental changes such as latency and resource limitations. Gagliolo and Schmidhuber proposed the Dynamic Algorithm Portfolio (DAP) approach, demonstrating through the multi-armed bandit problem that running multiple algorithms in parallel can yield optimal performance [8]. However, Jaggard et al. showed that heuristic algorithms in asynchronous environments may fail to converge, as explained by the Independence of Decisions (IoD) principle [9].

5. Optimization and Sequence Processing

Dynamic Programming (DP) is widely used to solve complex NLP problems by dividing them into smaller and repetitive subproblems. As analyzed by Zhang, DP improves computational efficiency through the optimal substructure property [10]. For model training, Makamo demonstrated that the Adam optimizer achieves faster and more accurate results than RMSprop and SGD on sequential datasets [11].

6. Large Language Models (LLM) and Serving Technologies

Running Large Language Models (LLMs) such as GPT-4 and LLaMA in distributed environments requires efficient memory management. The vLLM (PagedAttention) system proposed by Kwon et al. addresses KV-cache fragmentation using an operating-system–style paging mechanism, thereby improving inference throughput [12]. Hakim and Hasan demonstrated that systems such as PETALS enable collaborative inference of LLMs on mid-range user machines, effectively democratizing access to large models [13]. This approach enables the deployment of large models for low-resource languages, including Uzbek, in open distributed environments [14].

III. METHODOLOGY

The Adaptive Distributed NLP Processing (ADNLP) algorithm is based on a distributed computing methodology developed on Apache Spark to efficiently process large-scale text corpora. In this approach, NLP tasks are executed in parallel using adaptive allocation, dynamic partitioning, and GPU-aware scheduling mechanisms. The ADNLP architecture consists of the following main components:

– Corpus Analyzer — determines statistical characteristics of the input text corpus (number of tokens, sentence length, complexity level);

ISSN 3030-3907

–Dynamic Partition Engine — divides the text into adaptive segments based on length and complexity;
– Adaptive Scheduler — distributes tasks according to available node resources;
– Distributed Executors — run the BERT+BiLSTM model in parallel;
– Result Aggregator — merges distributed outputs and generates the final text.

In the ADNLP algorithm, data are divided into dynamic partitions, and each partition is sent to a separate computing node. This process can be expressed by the following mathematical model:

$$D=\{P_1,P_2,...,P_n\} \quad (1)$$

Here, P_i represents an adaptive segment. Punctuation restoration for each segment is performed as follows:

$$Y_i=f_{BERT+BiLSTM}(P_i) \quad (2)$$

After merging the results, the final output is obtained as follows:

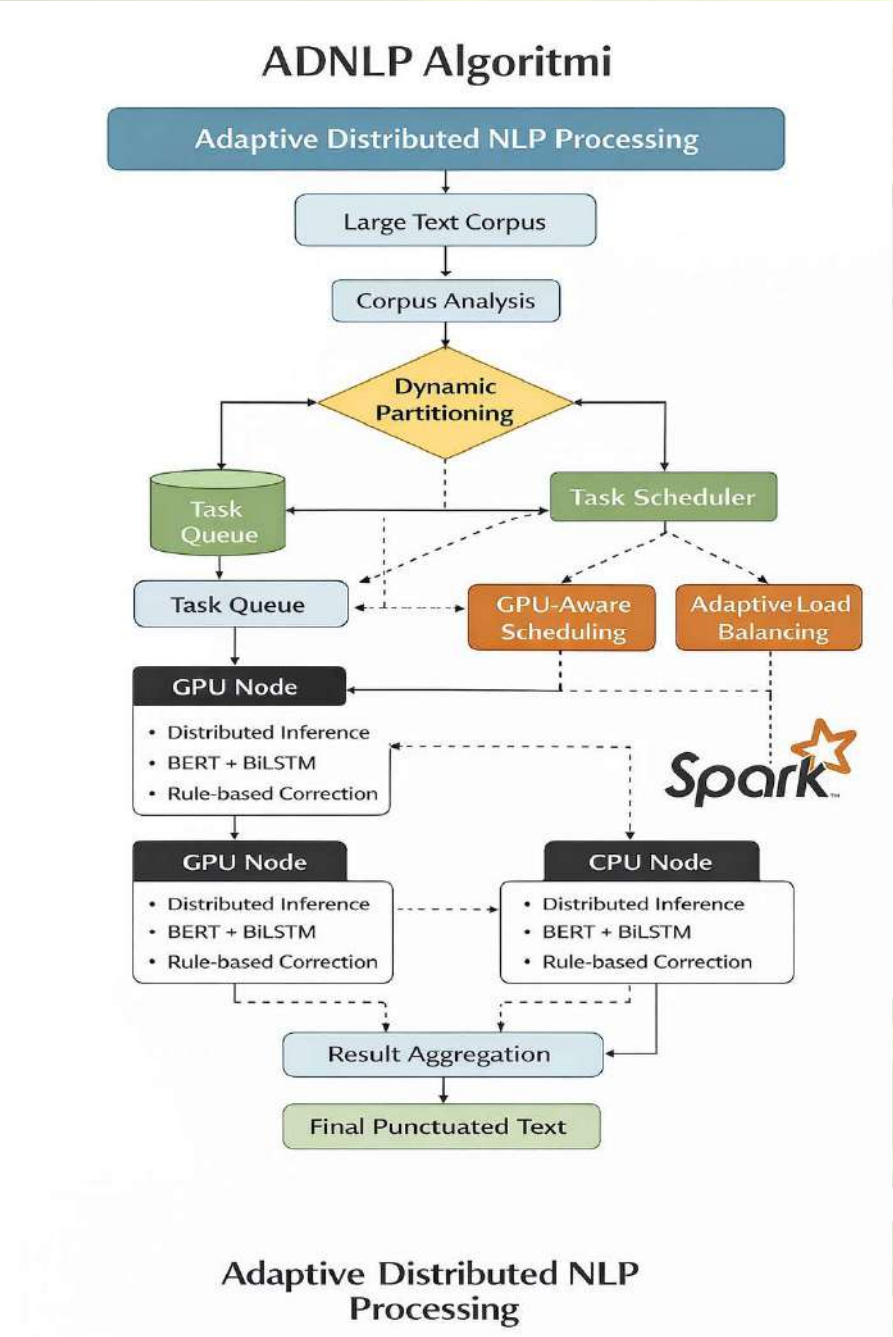
$$Y=\bigcup_{i=1}^{n} Y_i \quad (3)$$

The ADNLP methodology combines Spark's distributed computing mechanisms with GPU acceleration and adaptive load balancing. This approach enables parallel execution of NLP processes, including tokenization, embedding extraction, punctuation restoration, and rule-based post-editing.

Furthermore, the ADNLP algorithm is based on a pipeline architecture and operates through the following sequential stages:
– text analysis;
– dynamic partitioning;
– adaptive scheduling;
– BERT+BiLSTM inference;
– rule-based correction;
– result aggregation.

The proposed methodology ensures efficient utilization of computational resources and significantly reduces processing time when handling large-scale text corpora.

ISSN 3030-3907



Adaptive Distributed NLP Processing

IV. RESULT

A. Experimental Environment

- Sequential system: Intel i7-12700H, 32 GB RAM, 1 TB SSD.
- Spark cluster (Parallel): 5 computers, each equipped with GeForce RTX GPU, 32 GB RAM, and 1 TB SSD.

B. Dataset

Text corpora were prepared in sizes of 1 MB, 5 MB, 10 MB, 15 MB, 20 MB, 25 MB, 30 MB, and 35 MB.

C. Results Table

ISSN 3030-3907

Table 1

Presents the experimental results.

File Size (MB)	Sequential Execution Time	Distributed (5 nodes, Spark) Execution Time	ADNLP Algorithm (Proposed)
1 MB (46,938 words)	12 seconds	3.2 seconds	2.4 seconds
5 MB (234,605 words)	55 seconds	5.12 seconds	3.9 seconds
10 MB (480,658 words)	130 seconds	9.77 seconds	7.1 seconds
15 MB (704,070 words)	228.61 seconds	14.42 seconds	10.3 seconds
20 MB (938,762 words)	350.83 seconds	19.05 seconds	13.6 seconds
25 MB (1,173,562 words)	496.67 seconds	23.68 seconds	16.8 seconds
30 MB (1,408,165 words)	666.11 seconds	28.31 seconds	19.9 seconds
35 MB (1,642,860 words)	859.17 seconds	32.93 seconds	23.5 seconds

D. Analysis

Table 1 compares the efficiency of the proposed ADNLP algorithm with sequential and Spark-based distributed computing approaches. The results show that the ADNLP algorithm significantly reduces computation time through dynamic partitioning and adaptive load balancing mechanisms. The advantage of the algorithm becomes more pronounced as the data size increases. For a 35 MB text corpus, the ADNLP algorithm achieves approximately 36× speedup compared to sequential processing and about 1.4× improvement over Spark-based distributed computing.

ISSN 3030-3907

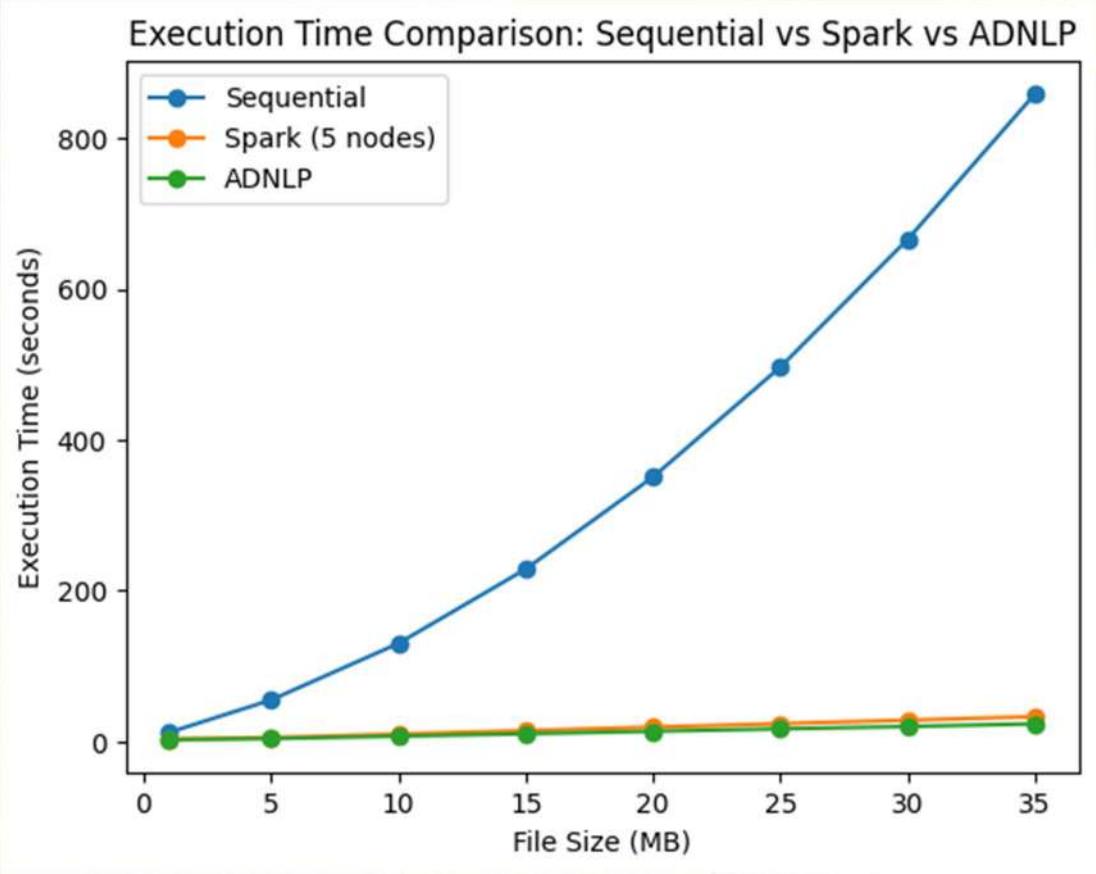


Figure 1. Comparative analysis graph of the experimental results

IV. CONCLUSION

This paper presented the ADNLP (Adaptive Distributed NLP Processing) algorithm for large-scale NLP tasks. By integrating distributed computing, adaptive scheduling, and GPU acceleration, the proposed approach significantly improves computational efficiency and scalability when processing large text corpora. The experimental results demonstrate that the ADNLP algorithm effectively reduces execution time compared to both sequential and Spark-based distributed approaches, while maintaining processing accuracy. The use of dynamic partitioning and adaptive load balancing enables efficient utilization of heterogeneous computing resources and minimizes idle time across cluster nodes.

Furthermore, the proposed methodology provides a flexible pipeline architecture suitable for large-scale punctuation restoration and other NLP applications. The algorithm can be extended to support additional NLP components such as token-level classification, sequence labeling, and multilingual processing.

Future work will focus on integrating CRF layers to improve sequence prediction accuracy, optimizing inter-node model communication, and reducing data transfer overhead. Additional improvements may include dynamic resource-aware scheduling, model caching mechanisms, and support for large language model (LLM) inference in distributed environments. These enhancements are

ISSN 3030-3907

expected to further improve the performance and scalability of the ADNLP framework for low-resource languages, including Uzbek.

REFERENCES

1. Wigan, M. R., & Clarke, R. (2013). Big Data's Big Unintended Consequences. *IEEE Computer Society*.
2. Omar, H. K., & Jumaa, A. K. (2022). Distributed big data analysis using spark parallel data processing. *BEEI*.
3. Andersson, L. (2016). Natural Language Processing In A Distributed Environment: A comparative performance analysis of Apache Spark and Hadoop MapReduce. *Diva-Portal*.
4. Ikramova, D. T., & Ochilova, M. (2026). Uzbek Agglutinative Morphology vs. English Analytical. *Shokh Library*.
5. Salaev, U. (2024). UzMorphAnalyser: A Morphological Analysis Model for the Uzbek Language Using Inflectional Endings. *MPASE2024*.
6. Abdurakhmonova, N., et al. (2025). Morpheme segmentation for Uzbek diversity domains. (Reference cited in Salaev & Matlatipov survey).
7. Walner, A. (2025). Ray vs Dask vs Apache Spark – Comparing DS & ML Engines. *Onehouse*.
8. Gagliolo, M., & Schmidhuber, J. (2006). Learning Dynamic Algorithm Portfolios. *IDSIA*.
9. Jaggard, A. D., Schapira, M., & Wright, R. N. (2011). Distributed Computing with Adaptive Heuristics.
10. Zhang, Y. (2024). A survey of dynamic programming algorithms. *Applied and Computational Engineering*.
11. Makamo, S. (2024). A Review of Optimization Algorithms in Deep Learning Models. *ResearchGate*.
12. Kwon, W., et al. (2023). Efficient Memory Management for LLM Serving with PagedAttention. *SOSP*.
13. Hakim, S. A., & Hasan, S. (2025). Explorative Study on Distributed Computing Techniques in LLMs.
14. Borzunov, A., et al. (2022). Petals: Collaborative inference and fine-tuning of large models.

ISSN 3030-3907