_____

COMPUTER METHODS IN LITERARY STUDIES: USING ARTIFICIAL INTELLIGENCE, BIG DATA, AND ALGORITHMIC APPROACHES FOR TEXT ANALYSIS

Kadirova M. X.
Lecturer, Department of Russian Literary Studies
Chirchik State Pedagogical University

Abstract

The article explores the transformative impact of computer-based methods on literary studies, with a focus on the combined use of artificial intelligence (AI), big data processing, and algorithmic modelling for in-depth analysis and interpretation of literary texts with unprecedented depth and scale.

Keywords: Digital literary analysis, machine learning, natural language processing, distant reading, stylometric profiling, thematic clustering, semantic embeddings, computational humanities.

КОМПЬЮТЕРНЫЕ МЕТОДЫ В ЛИТЕРАТУРОВЕДЕНИИ: ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА, БОЛЬШИХ ДАННЫХ И АЛГОРИТМИЧЕСКИХ ПОДХОДОВ ДЛЯ ИССЛЕДОВАНИЯ ТЕКСТОВ

Кадирова М.Х.
Преподаватель кафедры
«Русского литературоведения»
Чирчикского государственного педагогического университета

Аннотация:

В статье исследуется трансформирующее влияние компьютерных методов на литературоведение с акцентом на совместное применение искусственного интеллекта (ИИ), обработки больших данных и

_____

алгоритмического моделирования для детального анализа и интерпретации литературных текстов с беспрецедентной глубиной и масштабом.

Ключевые слова: цифровой анализ литературы, машинное обучение, обработка естественного языка, дистантное чтение, стилометрическое профилирование, тематическая кластеризация, семантические вложения, вычислительные гуманитарные науки.

Introduction

The digital revolution has profoundly altered the terrain of humanities research, transforming literary studies from predominantly qualitative, close-reading practices into a hybrid domain where macroscopic patterns emerge through computational lenses. Computational literary studies (CLS) stands at the confluence of literature, linguistics, informatics, and statistical modeling, employing artificial intelligence (AI), natural language processing (NLP), and expansive digital corpora to reveal latent structures, evolutionary trends, and stylistic signatures that often evade traditional scrutiny. In an age of information abundance, these tools empower researchers to interrogate millions of pages—from medieval manuscripts to social media fiction—yielding insights at scales previously unimaginable.

In Central Asia, where initiatives to digitize epic poetry, classical prose, and folk narratives are gaining momentum, AI-driven analytics offer invaluable avenues for cultural preservation and cross-temporal comparison. Uzbek students, increasingly exposed to programming fundamentals via integrated STEM-humanities tracks, are uniquely positioned to leverage open-source frameworks such as Python's NLTK, spaCy, or Hugging Face transformers. The region's multilingual literary ecosystem further necessitates sophisticated multilingual embeddings and transfer-learning paradigms, which modern AI architectures readily support. Beyond research, these methods invigorate pedagogy by fostering data-literate humanists capable of navigating both hermeneutic depth and empirical breadth.

The impetus for computational intervention arises from intractable questions in literary history: tracing genre diffusion, quantifying influence networks, or

_____

validating authorship amid textual variants. By converting interpretive hunches into testable hypotheses, CLS bridges subjective exegesis with reproducible evidence, enriching scholarly discourse while democratizing access to complex analytical workflows.

Thus, computational methodologies transcend mere technological augmentation; they constitute a foundational shift toward scalable, multimodal literary inquiry in a data-saturated world.

The urgency of AI adoption in literary analysis is underscored by the explosive growth of born-digital texts and the fragility of analog archives. Global benchmarks like the HathiTrust Digital Library and the Europeana Collections illustrate how algorithmic pipelines can map diachronic lexical shifts or sentiment trajectories across centuries. In Uzbekistan, digitization of Alisher Navoi's divans or Abdulla Qodiriy's novels creates fertile ground for applying topic modeling to chart ideological transitions from medieval mysticism to modernist realism. Students' prior exposure to English technical terminology—terms like "tokenization," "lemmatization," or "BERT"—streamlines onboarding, as these concepts function as international linguistic bridges.

Practical consolidation involves guided exercises: scripting frequency distributions, training naïve Bayes classifiers on authorial corpora, or visualizing character interaction graphs. Vocabulary selection prioritizes methodological transparency, reproducibility, and alignment with research ethics, favoring interpretable algorithms over opaque black-box systems.

Topic Modeling (LDA, BERTopic): Extracts probabilistic theme clusters from document collections, enabling macro-genre analysis.

Stylometry: Employs multivariate statistics on function words, n-grams, and syntactic markers to construct authorial DNA.

Sentiment Analysis: Deploys transformer-based models to trace emotional arcs across narrative acts, revealing affective architecture.

Network Analysis: Represents intertextual citations or character relationships as dynamic graphs, quantifying centrality and influence.

Word Embeddings (fastText, multilingual BERT): Generates contextual vector spaces to probe semantic evolution and cross-lingual analogies.

Sample Workflow for Literary Corpus Analysis:

_____

Using Python and the Gensim ecosystem, preprocess a bilingual Uzbek-Russian poetry corpus via tokenization, stop-word removal, and subword segmentation. Train a BERTopic model to surface latent motifs (e.g., "destiny," "homeland," "transience"); render interactive visualizations with pyLDAvis. For stylometry, calculate Burrows' Delta across suspected pseudonymous works; statistical significance validates or refutes attribution claims.

Explanation of foundational concepts:

• Corpus: A systematically curated textual dataset for computational interrogation.

• Tokenization: Segmenting raw text into minimal meaningful units (words, subwords).

• Lemmatization: Normalizing inflected forms to dictionary headwords.

• Vector Space Model: Encoding documents as high-dimensional numerical arrays for similarity metrics.

• Transformer Architecture: Attention-driven neural networks excelling in contextual representation.

• Embedding Layer: Learned dense vectors capturing semantic and syntactic regularities.

• Hyperparameter Tuning: Systematic optimization of model configurations for task-specific performance.

These instruments facilitate hypothesis testing, pattern visualization, and predictive modeling, converting literary scholarship into a collaborative, verifiable enterprise. In classroom settings, Jupyter Notebook-based labs enable students to replicate published studies or launch original micro-projects, cultivating computational thinking alongside critical interpretation.

Application to Russian Literary Studies

The application of computational methods to Russian literary studies offers a transformative lens through which to re-evaluate the rich tapestry of Russian literature, spanning from the medieval chronicles of Nestor to the modernist experiments of the Silver Age and beyond. Russian literary heritage, with its complex interplay of narrative styles, ideological shifts, and cultural symbolism, presents a fertile ground for AI-driven analysis. For instance, stylometric analysis

_____

can illuminate the authorship debates surrounding works attributed to Alexander Pushkin or the disputed texts of Nikolai Gogol, leveraging statistical patterns in lexicon and syntax to resolve long-standing scholarly disputes. Topic modeling applied to the vast corpus of 19th-century Russian novels—such as those by Leo Tolstoy and Fyodor Dostoevsky—can uncover thematic underpinnings like existential struggle, societal critique, or spiritual redemption, offering new interpretive frameworks. Sentiment analysis, meanwhile, can trace the emotional evolution in Anton Chekhov's short stories, revealing subtle shifts in tone that reflect his psychological depth.

In the Soviet and post-Soviet eras, computational tools enable the mapping of ideological influences across authors like Maxim Gorky and Boris Pasternak, highlighting how censorship and state narratives shaped literary production.

Network analysis can further elucidate intertextual relationships, such as the influence of Symbolism on early 20th-century poets like Alexander Blok or the dialogue between émigré writers and their homeland counterparts. The digitization of Russian archives, including the Pushkin House collections, amplifies these capabilities, allowing scholars to explore regional dialects, manuscript variants, and unpublished works with precision.

For Uzbek students studying Russian literature, these methods provide a bridge to engage with canonical texts while developing computational skills. Exercises might include annotating Dostoevsky's "Crime and Punishment" for moral dilemmas using NLP tools or constructing network graphs of character interactions in Tolstoy's "War and Peace." This interdisciplinary approach not only deepens literary understanding but also fosters a critical awareness of

These instruments facilitate hypothesis testing, pattern visualization, and predictive modeling, converting literary scholarship into a collaborative, verifiable enterprise. In classroom settings, Jupyter Notebook-based labs enable students to replicate published studies or launch original micro-projects, cultivating computational thinking alongside critical interpretation.

Augmentation strategies include virtual text-mining sandboxes on platforms like Google Colab, supporting real-time experimentation with multilingual datasets and pre-trained models.

_____

Conclusion

The fusion of AI, big data analytics, and algorithmic rigor within literary studies heralds a renaissance of the discipline, amplifying hermeneutic insight with empirical precision and global scalability. Far from supplanting humanistic intuition, computational approaches enhance it by illuminating macro-trends, quantifying stylistic evolution, and enabling cross-cultural synthesis in an increasingly digitized textual universe. Within the Uzbek academic ecosystem, such methodologies not only safeguard intangible heritage through advanced curation but also nurture a generation of scholars fluent in both poetic nuance and programming logic. By prioritizing open-source tools, ethical AI governance, and interdisciplinary training, CLS ensures inclusivity, transparency, and innovation. Ultimately, this convergence positions literary scholarship at the vanguard of knowledge production—where algorithms and imagination coalesce to decode the multifaceted tapestry of human narrative.

REFERENCES:

1. Jockers, M. L. Macroanalysis: Digital Methods and Literary History. University of Illinois Press, 2013.

2. Underwood, T. Distant Horizons: Digital Evidence and Literary Change. University of Chicago Press, 2019.

3. Algee-Hewitt, M., & McGurl, M. Between Canon and Corpus: Six Perspectives on 20th-Century Novels. Stanford Literary Lab, 2015.

4. Бочаров, С. Г. О поэтике Достоевского. Москва: Языки славянской культуры, 2001.

5. Лотман, Ю. М. Структура художественного текста. Санкт-Петербург: Искусство-СПБ, 2002.

6. Топоров, В. Н. Миф. Ритуал. Символ. Образ: Исследования в области мифопоэтических текстов. Москва: Прогресс-Традиция, 2004.