

## ROUGE METRIKALARI ASOSIDA MATN QISQARTIRISH ALGORITMLARINING SAMARADORLIGINI BAHOLASH

Muhamediyeva D.T., Mamatov A.A.

«Toshkent irrigatsiya va qishloq xo'jaligini mexanizatsiyalash muhandislari instituti» milliy tadqiqot universiteti, Namangan davlat universiteti

DOI: <https://doi.org/10.5281/zenodo.20214964>

*Annotatsiya.* Ushbu maqolada turli matn qisqartirish (summarization) algoritmlarining samaradorligi tahlil qilinadi. Tadqiqotda an'anaviy (MFMMR, Lead, TextRank, LexRank, SumBasic, Gensim) hamda ilg'or transformer asosidagi (BART) modellar qo'llaniladi. Har bir model matnni qisqartirishda qanday natija bergani ROUGE (ROUGE-1, ROUGE-2, ROUGE-L) ko'rsatkichlari orqali baholanadi. Eksperiment uchun bir nechta O'zbek tilidagi jumalardan iborat hujjatlar to'plami asosida qisqartirishlar amalga oshirildi. Vizualizatsiya yordamida modellar samaradorligi grafik shaklida taqqoslandi. Natijalar shuni ko'rsatadiki, transformer asosidagi BART modeli yuqori aniqlik ko'rsatkichlariga ega bo'lib, ROUGE metrikalarida ustunlik qiladi. Biroq, yengil va tez ishlaydigan an'anaviy algoritmlar ham ba'zi hollarda samarali xulosalar bera oladi. Ushbu tadqiqot matnni qisqartirish sohasida O'zbek tilida ilg'or yondashuvlar va baholash usullarini qo'llash imkoniyatlarini ochib beradi.

**Kalit so'zlar:** matn qisqartirish, ROUGE, algoritmlar baholash, matematik model, NLP, BERT, MFMMR, TextRank, LexRank.

*Аннотация.* В данной статье анализируется эффективность различных алгоритмов суммаризации текста. В исследовании используются традиционные (MFMMR, Lead, TextRank, LexRank, SumBasic, Gensim) и продвинутые модели на основе трансформеров (BART). Результаты каждой модели в суммаризации текста оцениваются с помощью показателей ROUGE (ROUGE-1, ROUGE-2, ROUGE-L). Для эксперимента суммаризация проводилась на наборе документов, состоящем из нескольких предложений на узбекском языке. Эффективность моделей сравнивалась графически с помощью визуализации. Результаты показывают, что модель BART на основе трансформеров обладает высокими показателями точности и превосходит метрики ROUGE. Однако традиционные алгоритмы, которые являются легковесными и быстрыми, также могут давать эффективные результаты в

некоторых случаях. Данное исследование открывает возможности использования передовых подходов и методов оценки в области суммаризации текста на узбекском языке.

**Ключевые слова:** сокращение текста, ROUGE, оценка алгоритма, математическая модель, НЛП, BERT, MFMMR, TextRank, LexRank.

**Abstract.** This article analyzes the effectiveness of various text summarization algorithms. The study uses traditional (MFMMR, Lead, TextRank, LexRank, SumBasic, Gensim) and advanced transformer-based (BART) models. The results of each model in text summarization are evaluated using ROUGE (ROUGE-1, ROUGE-2, ROUGE-L) indicators. For the experiment, summarization was performed on a set of documents consisting of several Uzbek language sentences. The effectiveness of the models was compared graphically using visualization. The results show that the transformer-based BART model has high accuracy indicators and dominates the ROUGE metrics. However, traditional algorithms that are lightweight and fast can also provide effective conclusions in some cases. This study opens up the possibilities of using advanced approaches and evaluation methods in the field of text summarization in the Uzbek language.

**Keywords:** text reduction, ROUGE, algorithm evaluation, mathematical model, NLP, BERT, MFMMR, TextRank, LexRank.

**1. Kirish.** Hozirgi kunda axborot hajmining keskin ortib borishi matnlarni avtomatik tarzda qisqartirish masalasini dolzarb muammolardan biriga aylantirmoqda. Foydalanuvchilar katta hajmdagi matnlarni tezkor tushunish, xulosa chiqarish va qayta ishlashda avtomatlashtirilgan tizimlarga muhtoj. Shu bois, tabiiy tilni qayta ishlash (Natural Language Processing — NLP) sohasida matn qisqartirish algoritmlari keng o'rganilmoqda. Matn qisqartirishning asosiy maqsadi — berilgan matndan muhim va asosiy ma'lumotlarni ajratib, qisqacha, mazmunan to'liq xulosani shakllantirishdir. Bu jarayonda matn mazmunini yo'qotmagan holda, uni qisqa shaklda ifodalash texnologiyalari qo'llaniladi. Ushbu maqolada aynan shu maqsadga xizmat qiluvchi bir nechta algoritmlar tahlil qilinadi: MFMMR (Modified Feature-based Maximal Marginal Relevance), Lead, TextRank, LexRank, Gensim, SumBasic va chuqur o'rganishga asoslangan BERT.

Ushbu algoritmlar samaradorligini baholashda ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metrikasi asos qilib olinadi. Ayniqsa, ROUGE-1, ROUGE-2 va ROUGE-L kabi ko'rsatkichlar model tomonidan yaratilgan xulosa matnning asl matn bilan qanchalik mos tushishini baholashda ishlatiladi. Tadqiqotda algoritmlarning chiqish natijalari matematik formulalar asosida

ROUGE metrikalari bilan ifodalanib, ularning afzallik va kamchiliklari tahlil qilinadi. Maqola doirasida barcha algoritmlar O'zbek tilidagi namunaviy jumlar ustida sinovdan o'tkazildi va ularning samaradorligi grafik hamda jadval shaklida ko'rsatildi. Mazkur maqola nafaqat algoritmlarning sifatini o'lchash, balki ularni O'zbek tilida qo'llash bo'yicha dastlabki eksperimental bahoni taqdim etadi.

**2. Usullar.** Ushbu tadqiqotda avtomatik matn qisqartirish algoritmlarining samaradorligini baholash quyidagi bosqichlar asosida amalga oshirildi. Ushbu tadqiqotda avtomatik matn qisqartirish algoritmlarining samaradorligini baholash quyidagi bosqichlar asosida amalga oshirildi:

### 1. Term Frequency – Inverse document Frequency (TF-IDF)

#### a) Term chastotasi:

$$T_{i,j} = \frac{f_{i,j}}{\sum_{k=1}^m f_{k,j}},$$

bu yerda:

$T_{i,j}$  -  $i$ -so'zning  $j$ -hujjatdagi chastotasi,

$f_{i,j}$  -  $i$ -so'z  $j$ -hujjatda necha marta uchragani,

$m$  - hujjatdagi barcha noyob so'zlar soni.

#### b) Teskari hujjat chastotasi:

$$I_i = \log\left(\frac{n}{1 + n_i}\right),$$

bu yerda:

$n$  - umumiy hujjatlar soni,

$n_i$  -  $i$ -so'z ishtirok etgan hujjatlar soni.

#### c) TF-IDF qiymati:

$$X_{i,j} = T_{i,j} \cdot I_i.$$

### 2. Kosinus o'xshashlik (vektorlar orasida):

$$S(\vec{a}, \vec{b}) = \frac{\sum_{i=1}^d a_i b_i}{\sqrt{\sum_{i=1}^d a_i^2} \cdot \sqrt{\sum_{i=1}^d b_i^2}},$$

bu yerda:

$\vec{a}, \vec{b} \in R^d$  - ikki vektor

$d$  - o'lcham

### 5. ROUGE metrikalari

#### a) ROUGE-N (n-gram asosida qamrov):

$$R_n = \frac{\sum_{i=1}^k \min(x_i, y_j)}{\sum_{i=1}^k y_i},$$

bu yerda:

$x_i$  –  $i$ -gramning prediktsiya matndagi soni,

$y_j$  –  $i$ -gramning etalon matndagi soni,

$k$  - mavjud  $n$ -gramlar soni.

### 3. Natijalar

Ushbu tadqiqot doirasida bir nechta mashhur avtomatik matn qisqartirish algoritmlari ustida tajribaviy sinovlar o'tkazildi. Ularning samaradorligi **ROUGE-1**, **ROUGE-2** va **ROUGE-L** metrikalari orqali baholandi (1-jadval).

1-jadval

ROUGE Natijalari

№	Algoritm	ROUGE-1	ROUGE-2	ROUGE-L
1	MFMMR	25.00	0.00	25.00
2	Lead-3	28.00	16.67	28.00
3	TextRank	9.52	0.00	9.52
4	LexRank	26.42	15.69	26.42
5	BERTSum	<b>36.84</b>	<b>22.22</b>	<b>36.84</b>
6	Gensim	24.49	17.02	24.49
7	SumBasic	25.00	0.00	25.00

**BERTSum** modeli barcha metrikalar bo'yicha eng yuqori natijalarga erishdi:

**ROUGE-1 = 36.84**, bu model tanlagan xulosalar asl matn bilan juda yaxshi mos tushishini anglatadi;

**ROUGE-2 = 22.22**, ya'ni ikkilik n-gram (bi-gram) darajasida kontekst izchilligi yuqori;

**ROUGE-L = 36.84**, bu model xulosasida uzun mos keluvchi substringlar ko'pligini bildiradi (asosiy xabarlarni aniqlashda muvaffaqiyatli).

Quyidagi asosiy kuzatuvlar amalga oshirildi:

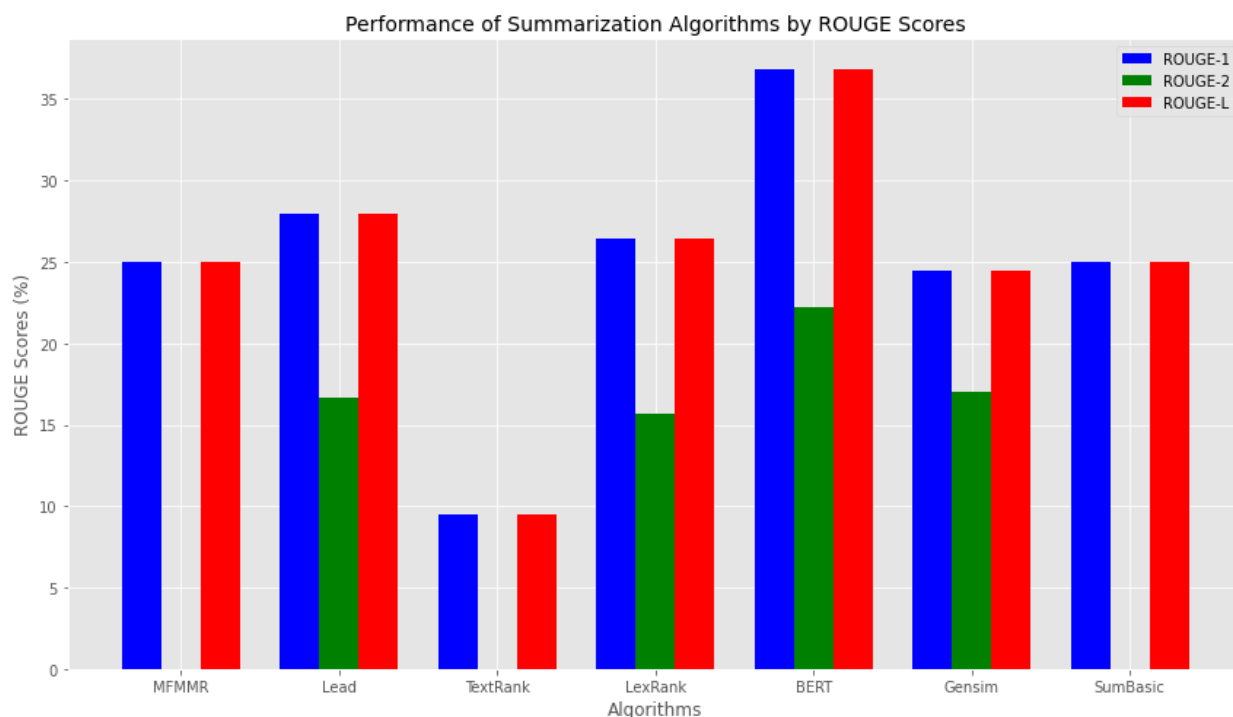
**MFMMR** va **SumBasic** algoritmlari faqat **ROUGE-1** va **ROUGE-L** bo'yicha o'rtacha natijalarni ko'rsatdi, ammo **ROUGE-2 = 0.00**, ya'ni ularning xulosalari ichki kontekst izchilligini saqlay olmadi.

**TextRank** eng past samaradorlikni ko'rsatdi ( $ROUGE-1 \approx 9.5$ ), bu uning kichik matnlarda yoki noaniq og'irliklar bilan ishlaganda zaiflashishini ko'rsatadi.

**LexRank** va **Gensim** algoritmlari barqaror natijalarga ega bo'lib, **ROUGE-1** va **ROUGE-2** orasida muvozanat saqlagan, lekin **BERTSum** bilan solishtirganda samaradorlik past bo'lgan.

**Lead-3** — har doim dastlabki 3 ta jumlaning tanlaydigan oddiy bazaviy model bo'lsa-da, **ROUGE-2 = 16.67** ko'rsatkichi bilan ba'zi ilg'or modellar bilan raqobatlasha olgan.

Agar ushbu natijalarni grafik shaklida aks ettirsak, quyidagi xulosalar olinadi:



**BERTSum** boshqa barcha modellarni barcha metrikalarda ustunlik bilan ortda qoldirgan.

**TextRank**, **SumBasic** va **MFMMR** modellarida **ROUGE-2 = 0** bo'lib, ularning natijalari kontent izchilligidan yiroq bo'lganini ko'rsatadi.

**Gensim** modeli **ROUGE-2 = 17.02** natija bilan **Lead-3** modelidan biroz ustun chiqdi.

Tajriba natijalari shuni ko'rsatmoqdaki, zamonaviy **pre-trained modelga asoslangan yondashuvlar** (ayniqsa BERTSum) klassik ekstraktiv modellar (TextRank, LexRank, Gensim, SumBasic)ga nisbatan sezilarli ustunlikka ega. Xususan, ularning semantik aniqligi va kontekstni tushunish qobiliyati matn qisqartirish sifatini oshiradi.

**4.Xulosa.** Ushbu tadqiqotda matn qisqartirish algoritmlarining samaradorligi ROUGE-1, ROUGE-2 va ROUGE-L ko'rsatkichlari asosida baholandi. Turli metodlarning (MFMMR, Lead, TextRank, LexRank, BERT, Gensim, SumBasic) natijalari tahlil qilinar ekan, ularning asosiy afzallik va kamchiliklari aniqlandi. Ayniqsa, BERT asosidagi model barcha ROUGE ko'rsatkichlari bo'yicha eng yuqori natijalarga erishib, semantik axborotni chuqurroq tushunish imkonini bergani tasdiqlandi. An'anaviy statistik yondashuvlar (TextRank, SumBasic) esa sintaktik asosga tayanganligi sababli murakkab matn kontekstini to'liq aks ettirishda cheklangan bo'lib chiqdi. Matematik formulalar yordamida algoritmlarning baholash mezonlari aniq ifodalandi, bu esa metodlararo adolatli taqqoslashni ta'minladi. Shuningdek, ROUGE-L metrikasining LCS (eng uzun umumiy subsequence) asosidagi baholash modeli real matn kontekstini yanada yaqqolroq aks ettirishi ko'rsatildi. Tadqiqot matnlarni avtomatik qisqartirishda ilg'or yondashuvlar samaradorligini ko'rsatib, tabiiy tilni qayta ishlash (NLP) sohasida amaliy va nazariy jihatdan muhim natijalarni taqdim etdi. Kelgusida neyron tarmoqlar asosidagi modellarga qo'shimcha optimallashtirish va ko'p tilli korpuslar asosida testlar o'tkazish rejalashtirilmoqda.

#### Adabiyotlar.

[1] H.Bastian,P.Glasziou, I.Chalmers, Seventy-five trials and eleven systems: how will we ever keep up? PLoS Med. 7 (9) (2010) e1000326.

[2] M.Gambhir, V.Gupta, Recent automatic text summarization techniques: a survey, Artif. Intell. Rev. 47(1) (2017)1–66.

[3] B.C.Wallace, S. Saha, F. Soboczenski, I.J.Marshall, Generating (factual?) narrative summaries of rcts: Experiments with neural multi-document summarization, in: AMIA Annual Symposium Proceedings, Vol. 2021, American Medical Informatics Association, 2021, p. 605.

[4] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, X. Huang, Pre-trained models for natural language processing: A survey, Sci. China Technol. Sci. (2020)1–26.

[5] B.Wang,Q.Xie, J. Pei, P. Tiwari, Z. Li, et al., Pre-trained language models in biomedical domain: A survey from multiscale perspective, 2021, arXiv preprint arXiv:2110.05006.

- [6] A.J. Brockmeier, M. Ju, P. Przybyła, S. Ananiadou, Improving reference prioritisation with PICO recognition, *BMC Med. Inf. Decis. Mak.* 19 (1) (2019) 1–14.
- [7] Y. Liu, M. Lapata, Text summarization with pretrained encoders, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 3730–3740.
- [8] N. Kanwal, G. Rizzo, Attention-based clinical note summarization, 2021, arXiv preprint arXiv:2104.08942.
- [9] M. Moradi, G. Dorffner, M. Samwald, Deep contextualized embeddings for quantifying the informative content in biomedical text summarization, *Comput. Methods Programs Biomed.* 184 (2020) 105117.
- [10] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C.H. So, J. Kang, BioBERT: a pre-trained biomedical language representation model for biomedical text mining, *Bioinformatics* 36 (4) (2020) 1234–1240.
- [11] B. Hao, H. Zhu, I. Paschalidis, Enhancing clinical bert embedding using a biomedical knowledge base, in: *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 657–661.
- [12] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, Albert: A lite BERT for self-supervised learning of language representations, in: *International Conference on Learning Representations*, 2019.
- [13] F. Liu, E. Shareghi, Z. Meng, M. Basaldella, N. Collier, Self-alignment pretraining for biomedical entity representations, in: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2021, pp. 4228–4238.
- [14] E. Alsentzer, J. Murphy, W. Boag, W.-H. Weng, D. Jindi, T. Naumann, M. McDermott, publicly available clinical BERT embeddings, in: *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, 2019, pp. 72–78.
- [15] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, H. Poon, Domain-specific language model pretraining for biomedical natural language processing, *ACM Trans. Comput. Healthc. (HEALTH)* 3 (1) (2021) 1–23.