

УДК 004.89
МРНТИ 28.23.29

<https://doi.org/10.55452/1998-6688-2025-22-1-36-43>

^{1*}**Нұрсана С.Т.**,
магистрант, ORCID ID: 0009-0005-7104-4646,
*e-mail: samal200174@gmail.com
¹**Уалиева И.М.**,
к.ф.-м.н., ассоциированный профессор, ORCID ID: 0000-0003-3853-8896,
e-mail: ualiyeva.irina@kaznu.kz

¹Казахский национальный университет им. аль-Фараби, г. Алматы, Казахстан

ПРИМЕНЕНИЕ МЕТРИКИ BLEU И SARI В ОЦЕНКЕ УПРОЩЕННЫХ ТЕКСТОВ НА КАЗАХСКОМ ЯЗЫКЕ: АНАЛИЗ И ЭФФЕКТИВНОСТЬ

Аннотация

В данной статье рассматривается методика оценки качества упрощенных текстов на казахском языке с использованием метрик BLEU и SARI. Упрощение текстов является важным аспектом для обеспечения доступности информации и облегчения процесса обучения на казахском языке. Метрика BLEU, основанная на сравнении n-грамм перевода и эталона, широко используется для оценки качества машинного перевода, но не учитывает контекст входного текста. Метрика SARI, специально разработанная для оценки упрощения текста, учитывает изменения в семантике и демонстрирует более высокую корреляцию с оценками человека. В рамках исследования были применены алгоритмы замены сложных слов простыми синонимами и алгоритмы замены или удаления сложных фраз. Результаты анализа показали, что метрика SARI более чувствительна к изменениям, внесенным в упрощенные тексты, по сравнению с BLEU. Таким образом, комбинированное использование метрик BLEU и SARI обеспечивает всестороннюю и точную оценку качества упрощенных текстов на казахском языке.

Ключевые слова: упрощение текстов, казахский язык, метрики оценки, BLEU, SARI, качество текстов, машинный перевод, автоматическое упрощение текста, алгоритмы упрощения.

Введение

Цель упрощения текста заключается в переработке исходного текста таким образом, чтобы выходной текст стал более удобочитаемым. Упрощение текста находит применение в снижении сложности входных данных для обработки естественного языка [1] и в предоставлении помощи в чтении людям с ограниченными языковыми навыками [2]. Упрощение текстов также способствует устранению языковых барьеров и повышению уровня вовлеченности широких слоев населения в различные сферы жизни, включая образование, медиа и цифровую среду. Упрощение текстов на казахском языке также играет важную роль в обеспечении доступности информации и упрощении процесса обучения и общения на этом языке.

Однако отсутствие эффективных методов оценки качества упрощенных текстов затрудняет разработку инструментов и подходов для упрощения текста [3]. В результате возникает необходимость в создании и адаптации метрик, которые могли бы адекватно оценивать качество упрощенных текстов и обеспечивать высокую доступность и точность информации. Такой подход позволит не только улучшить существующие методы упрощения [4], но и ускорить их внедрение в практическую деятельность, что в итоге приведет к более широкому использованию казахского языка в различных контекстах.

Метрика BLEU (Bilingual Evaluation Understudy) является одной из наиболее широко используемых метрик для оценки качества машинного перевода [5]. Она основана на сравнении n-грамм перевода с эталонными переводами и позволяет получить количественную оценку

сходства между ними. Однако BLEU не учитывает контекст входного текста, что может приводить к недостаточной точности при оценке качества упрощенных текстов.

Метрика SARI (System output Against References and against the Input sentence) была специально разработана для оценки качества автоматического упрощения текста [6]. Она учитывает изменения в семантике и структуре текста, а также демонстрирует более высокую корреляцию с оценками человека. Это делает SARI более подходящей для задач упрощения текста, так как она позволяет оценивать как правильность добавленных слов, так и точность удалений и сохранений важных элементов текста.

В рамках данного исследования были применены два алгоритма упрощения текста на казахском языке: замена сложных слов простыми синонимами и замена или удаление сложных фраз. Эти алгоритмы позволяют снизить сложность текста, сохраняя его смысловую нагрузку и улучшая доступность информации.

Основной целью данного исследования является применение метрик BLEU и SARI для оценки качества упрощенных текстов на казахском языке и проведение сравнительного анализа их эффективности. Результаты анализа показали, что метрика SARI более чувствительна к изменениям, внесенным в упрощенные тексты, чем BLEU, что подчеркивает важность использования специализированных метрик для оценки упрощения текста.

Постановка задачи

Основной задачей данного исследования является применение существующих метрик BLEU и SARI для оценки качества упрощенных текстов на казахском языке. В рамках данной задачи необходимо решить следующие подзадачи:

1. Анализ применимости метрик BLEU и SARI для казахского языка
2. Применение алгоритмов упрощения текста
3. Применение метрик BLEU и SARI для оценки качества упрощенных текстов
4. Сравнительный анализ результатов

Выводы и рекомендации

Данная работа направлена на создание комплексного подхода к оценке качества упрощенных текстов на казахском языке, что будет способствовать повышению доступности информации и улучшению образовательного процесса. Комбинированное использование метрик BLEU и SARI обеспечивает всестороннюю и точную оценку качества упрощенных текстов, учитывая как лексические, так и семантические аспекты упрощения, что позволяет выявить особенности каждого метода и их применимость к конкретным случаям [7].

Материалы и методы

Для оценки качества упрощенных текстов на казахском языке в данном исследовании используются две основные метрики: BLEU и SARI. Эти метрики позволяют количественно оценить сходство между упрощенным текстом и эталонным текстом, а также учитывать изменения, внесенные в текст для его упрощения.

Метрика BLEU была разработана для автоматической оценки качества машинного перевода. Она основывается на сравнении n-грамм переведенного текста с эталонным текстом и вычисляет точность совпадения этих n-грамм. Основные аспекты метрики BLEU включают:

- ♦ N-граммы: последовательности из n слов, используемые для сравнения;
- ♦ Precision (точность): отношение числа совпадающих n-грамм в переведенном тексте к общему числу n-грамм в этом тексте;
- ♦ Brevity Penalty (штраф за краткость): штраф за слишком короткий перевод, вводимый в случае, если переведенный текст короче эталонного.

BLEU широко используется для оценки машинного перевода благодаря своей простоте и эффективности, но у него есть свои ограничения [8]. В частности, BLEU не учитывает контекст входного текста и не всегда отражает семантические изменения.

Формула BLEU:

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N \omega_n \log(p_n)\right) \quad (1)$$

где:

$$\diamond BP = \begin{cases} 1, & \text{если } c > r \\ e^{1 - \frac{r}{c}}, & \text{если } c \leq r \end{cases}$$

- ♦ c – длина переведенного текста.
- ♦ r – длина эталонного текста.
- ♦ p_n – точность n -грамм.
- ♦ ω_n – вес для n -грамм (обычно все веса равны).

Метрика SARI была специально разработана для оценки качества автоматического упрощения текста. В отличие от BLEU, она учитывает изменения, внесенные в текст, и оценивает их полезность и правильность на основе сравнения с эталонными текстами и исходным текстом.

Основные аспекты метрики SARI:

- ♦ Additions (добавления): насколько новые слова, добавленные в упрощенный текст, полезны;
- ♦ Deletions (удаления): насколько правильно удалены сложные слова или фразы;
- ♦ Keeps (сохранения): насколько правильно сохранены важные слова и фразы из исходного текста.

Формула SARI:

$$SARI = \frac{1}{3} (Precision_{Add} + Precision_{Delete} + F1_{Keep}) \quad (2)$$

где:

$Precision_{Add}$: точность добавленных слов.

$F1_{Keep}$: точность удаленных слов.

$Precision_{Delete}$: F1-мера для сохраненных слов, которая учитывает как точность, так и полноту.

SARI демонстрирует более высокую корреляцию с оценками человека по сравнению с BLEU, что делает ее более подходящей для задач упрощения текста [9]. Она позволяет оценивать как правильность добавленных слов, так и точность удалений и сохранений важных элементов текста, что особенно важно для задач упрощения.

Алгоритмы упрощения текстов. В рамках данного исследования были применены два основных алгоритма упрощения текста на казахском языке: алгоритм замены сложных слов простыми синонимами и алгоритм удаления или замены сложных фраз более простыми выражениями.

Алгоритм замены сложных слов простыми синонимами направлен на замену сложных или редко используемых слов их более простыми и понятными синонимами. Для реализации алгоритма создается словарь, в котором каждому сложному слову соответствует его упрощенный синоним. Примерный процесс работы алгоритма следующий:

1. Создание словаря синонимов

- ♦ Составляется список сложных слов и их упрощенных аналогов.

♦ Пример: «құралды» заменяется на «жасалған».

2. Замена слов в тексте

♦ Текст анализируется, и каждое сложное слово заменяется на соответствующий синоним из словаря.

Пример: «Қазақ халқының дәстүрлі азық-түлік өнімдері басқа да көшпенді халықтар сияқты ет пен сүт тағамдарынан, сондай-ақ, өсімдік және аңшылық және балық өнімдерінен құралды» -> «Қазақ халқының дәстүрлі тағамдары ет, сүт, өсімдік, аңшылық және балық өнімдерінен жасалған».

Алгоритм удаления или замены сложных фраз более простыми выражениями направлен на упрощение сложных фраз и предложений, делая их более понятными для широкой аудитории. Основные этапы работы алгоритма включают:

1. Создание списка сложных фраз и их упрощенных версий

♦ Составляется список сложных фраз и их более простых аналогов.

♦ Пример: «материалдық мәдениеттің ерекше маңызды элементі бола отырып» -> «ерекше маңызды элементі болып».

2. Анализ текста и замена фраз

♦ Текст анализируется на наличие сложных фраз, которые заменяются на их упрощенные версии.

♦ Пример: «Ұлттық тағам түрлері материалдық мәдениеттің ерекше маңызды элементі бола отырып, ұлттық формасы мен дәстүрінің біршама тұрақтылығымен айрықшаланады». -> «Ұлттық тағам түрлері ұлттық формасы мен дәстүрінің тұрақтылығымен ерекшеленеді».

Эти алгоритмы позволяют уменьшить сложность текста, сохраняя его смысловую нагрузку и улучшая доступность информации.

Таблица 1 – Примеры упрощенных текстов [10]

Исходный текст	Упрощенный текст
Ұлттық тағам түрлері материалдық мәдениеттің ерекше маңызды элементі бола отырып, ұлттық формасы мен дәстүрінің біршама тұрақтылығымен айрықшаланады. Қазақ халқының дәстүрлі азық-түлік өнімдері басқа да көшпенді халықтар сияқты ет пен сүт тағамдарынан, сондай-ақ, өсімдік және аңшылық және балық өнімдерінен құралды.	Ұлттық тағам түрлері ұлттық формасы мен дәстүрінің тұрақтылығымен ерекшеленеді. Қазақ халқының дәстүрлі тағамдары ет, сүт, өсімдік, аңшылық және балық өнімдерінен құралған.
Бауырсақ – ұннан жасалатын қазақтың ұлттық тағамдары ішінде жалпы жұртшылыққа танымал түрлерінің бірі. Бауырсақты майға қуырып пісіреді. Ол үшін әуелі арнайы ашытылған қамырды әртүрлі көлемде және әртүрлі қалыпта үзіп жасайды. Мысалы, дөңгелек бауырсақ, сопақ бауырсақ, қытырлақ ши бауырсақ, жалпақ әрі үлкен бауырсақ секілді бауырсақтың түрлерін қамырдан кесіп алған кезде дайындайды.	Бауырсақ – қазақтың ұлттық тағамдарының бірі, майға қуырып пісіреді. Ол әртүрлі көлемде және қалыпта әртүрлі қамырды ашып жасайды, мысалы, дөңгелек, сопақ, қытырлақ ши, жалпақ әрі үлкен бауырсақ.
Дәнді дақылдардың ішінде жиі пайдаланылатын түрлерінің бірі – тары. Қазақстанның кей өңірлерінде қуырылып, келіге түйіліп, қабыршықтарынан тазартылған тарыны сөк деп, тек шикі күйінде ғана тары деп атайды. Сөкті сүтке салып та, диірменге тартып талқан қылып та жеуге болады. Сондай-ақ сөкке қант, сары май қосып жент те жасайды. Сөктен жасалған тағам өте құнарлы, тойымды тағам саналады. Ал оған айран құйып, қайнатып жасаған тағам тары көже деп аталады.	Дәнді дақылдардың ішінде жиі пайдаланылатын түрлерінің бірі – тары. Бұл Қазақстанда кей өңірлерде қуырылып, келіге түйіліп, қабыршықтарынан тазартылған тарады. Сөкке сүт салып, диірменге тартып талқан қылып та жеуге болады.
Қазақ халқы ыдыс-аяқтар мен үй ішілік тұрмыстық заттарды негізінен тері мен ағаштан және үй жануарларының сүйегінен жасаған. Алуан түрлі осынау заттардың арғы бастауын Еуразия даласындағы көшпенділер әлемінің материалдық мәдениетінен кездестіруге болады. Үй ішілік тұрмыстық заттар ішінде теріден жасалған әртүрлі мөлшердегі ыдыстар көптеп кездеседі.	Қазақ халқының ыдыс-аяқтары мен үй ішілік тұрмыстық заттары тері мен ағаштан, сондай-ақ үй жануарларының сүйегінен жасалған. Бұл осынау заттарын Еуразия даласындағы көшпенділер әлемінің материалдық мәдениетімен қарым-қатынасын сақтайды.
Айран дайындау үшін піскен сүтке ұйытқы салып, араластырып бір-екі сағат аузын жауып қояды. Ыстықта тез бұзылатын айранды мұздатқышқа қойып қана сақтауға болады. Бұрынғы кездері терең құдыққа салып тастаған. Жаз мезгілінде сусын ретінде ішу үшін оған аздап су немесе сүт араластырып шалап дайындайды.	Айран дайындау үшін сүтке ұйытқы салып, араластырып бір-екі сағат аузын жауып қояды. Ыстықта мұздатқышқа қойып сақтайды. Жаз мезгілінде ішу үшін аздап су немесе сүт араластырып шалап дайындайды.

Результаты и обсуждение

Для проведения оценки были использованы тексты, разделенные на три категории:

- ◆ Исходный текст
- ◆ Упрощенный текст
- ◆ Эталонный текст

Эти данные были загружены из CSV файла. Каждая строка файла содержала один набор текстов (исходный, упрощенный и эталонный), что позволяло автоматизировать процесс оценки.

Для расчета BLEU метрики использовалась библиотека nltk. BLEU измеряет точность совпадения n-грамм между упрощенным и эталонным текстами:

```
from nltk.translate.bleu_score import sentence_bleu, SmoothingFunction

def calculate_bleu(reference, hypothesis):
    reference_tokens = [reference.split()]
    hypothesis_tokens = hypothesis.split()
    smoothie = SmoothingFunction().method4
    bleu_score = sentence_bleu(reference_tokens, hypothesis_tokens, smoothing_function=smoothie)
    return bleu_score
```

Рисунок 1 – Расчет метрики BLEU

Для расчета SARI метрики была реализована специальная функция, учитывающая три аспекта: добавление новых слов, удаление сложных слов и сохранение важных слов из исходного текста:

```
def calculate_sari(source, simplified, references):
    source_tokens = source.split()
    simplified_tokens = simplified.split()
    reference_tokens = [ref.split() for ref in references]

    num_ref = len(reference_tokens)

    add_total, add_correct = 0, 0
    keep_total, keep_correct = 0, 0
    del_total, del_correct = 0, 0

    for word in simplified_tokens:
        if word not in source_tokens:
            add_total += 1
            if any(word in ref for ref in reference_tokens):
                add_correct += 1

    for word in source_tokens:
        if word not in simplified_tokens:
            del_total += 1
            if all(word not in ref for ref in reference_tokens):
                del_correct += 1

    for word in source_tokens:
        keep_total += 1
        if word in simplified_tokens and all(word in ref for ref in reference_tokens):
            keep_correct += 1

    add_precision = add_correct / add_total if add_total > 0 else 0
    del_precision = del_correct / del_total if del_total > 0 else 0
    keep_precision = keep_correct / keep_total if keep_total > 0 else 0

    add_recall = add_correct / num_ref
    del_recall = del_correct / num_ref
    keep_recall = keep_correct / num_ref

    add_f1 = 2 * (add_precision * add_recall) / (add_precision + add_recall) if (add_precision + add_recall) > 0 else 0
    del_f1 = 2 * (del_precision * del_recall) / (del_precision + del_recall) if (del_precision + del_recall) > 0 else 0
    keep_f1 = 2 * (keep_precision * keep_recall) / (keep_precision + keep_recall) if (keep_precision + keep_recall) > 0 else 0

    sari_score = (add_f1 + del_f1 + keep_f1) / 3
    return sari_score
```

Рисунок 2 – Расчет метрики SARI

Для каждого набора текстов (исходный, упрощенный, эталонный) рассчитывались значения BLEU и SARI. Полученные результаты:

(Исходный текст	Упрощенный текст	Эталонный текст	BLEU	SARI
0	Исходный 0	Упрощенный 0	Эталонный 0	0.604836	0.599231
1	Исходный 1	Упрощенный 1	Эталонный 1	0.573087	0.648968
2	Исходный 2	Упрощенный 2	Эталонный 2	0.612384	0.652864
3	Исходный 3	Упрощенный 3	Эталонный 3	0.656151	0.629886
4	Исходный 4	Упрощенный 4	Эталонный 4	0.568292	0.661936
..
95	Исходный 95	Упрощенный 95	Эталонный 95	0.506824	0.689266
96	Исходный 96	Упрощенный 96	Эталонный 96	0.594806	0.625807
97	Исходный 97	Упрощенный 97	Эталонный 97	0.593053	0.677686
98	Исходный 98	Упрощенный 98	Эталонный 98	0.580256	0.672910
99	Исходный 99	Упрощенный 99	Эталонный 99	0.568271	0.612851

Рисунок 3 – Результат экспериментов

Одной из основных трудностей было адаптировать существующие метрики для казахского языка. Казахский язык имеет свою специфическую грамматику и лексику, что потребовало внесения некоторых изменений в алгоритмы оценки. Например, для расчета BLEU необходимо было учитывать особенности казахских словоформ и их склонения. Для адекватного сравнения упрощенных текстов с эталонными потребовалось собрать достаточное количество эталонных текстов. Это оказалось непростой задачей, так как эталонные тексты должны были быть грамотно написаны и соответствовать уровню упрощенных текстов.

Полученные результаты показали среднюю оценку BLEU на уровне 0.58 и SARI на уровне 0.67 при анализе 100 предложений. Эти цифры свидетельствуют о достаточно высоком уровне сходства упрощенных текстов с эталонными версиями, что подтверждает эффективность процесса упрощения. Эти результаты согласуются с выводами ряда исследований, проведенных в области автоматического упрощения текста, указывая на более высокую чувствительность метрики SARI к изменениям, внесенным в упрощенные тексты. Однако отсутствие эффективных методов оценки качества упрощенных текстов затрудняет оценку эффективности процесса упрощения и разработку соответствующих инструментов и подходов [11].

Заключение

В данном исследовании были применены метрики BLEU и SARI для оценки качества упрощенных текстов на казахском языке. Эти метрики позволили провести всесторонний анализ текстов, выявить эффективность алгоритмов упрощения и определить наиболее подходящие подходы для повышения доступности информации.

Результаты показали, что применение алгоритмов замены сложных слов простыми синонимами и замены или удаления сложных фраз позволяет значительно снизить сложность текста, сохраняя при этом его смысловую нагрузку.

Метрика BLEU, широко используемая для оценки качества машинного перевода, показала свою полезность для оценки упрощенных текстов. Однако из-за ее ограниченной способности учитывать контекст входного текста, метрика BLEU не всегда отражает семантические изменения, внесенные в текст. В то же время метрика SARI, специально разработанная для оценки автоматического упрощения текста, продемонстрировала высокую чувствительность к изменениям в тексте, что делает ее более подходящей для задач упрощения.

Средние оценки BLEU (0.58) и SARI (0.67) свидетельствуют о достаточном уровне сходства упрощенных текстов с эталонными версиями. Эти результаты подтверждают, что пред-

ложенные алгоритмы упрощения успешно справляются с задачей, делая тексты более доступными без существенной потери информации.

Комбинированное использование метрик BLEU и SARI позволяет получить более полное представление о качестве упрощения текста. Метрика BLEU обеспечивает оценку лексического сходства, тогда как SARI оценивает структурные и семантические изменения. Таким образом, использование обеих метрик позволяет выявить сильные и слабые стороны различных подходов к упрощению текста.

Информация о финансировании. Работа проводится в рамках ГФ AP19680575 «Машинное обучение для полуавтоматического формирования/генерации образовательного контента».

REFERENCES

- 1 Jiang, Chao, Mounica Maddela, Wuwei Lan, Yang Zhong, Wei Xu. *Comput. Res. Repos.*, 2020. <https://doi.org/10.48550/arXiv.2005.02324>
- 2 Lindstrom, Jennifer H. *Teaching Exceptional Children*, 2019, vol. 51, pp. 189–200. <https://doi.org/10.1177/0040059918763712>.
- 3 Xu W., Callison-Burch C., Napoles C. *Transactions of the Association for Computational Linguistics*, 2015, vol. 3, pp. 283–297. https://doi.org/10.1162/tacl_a_00139.
- 4 Suha S. Al-Thanyyan, Aqil M. Azmi. *ACM Comput. Surv.*, 2021, 36 p. <https://doi.org/10.1145/3442695>.
- 5 Papineni K., Roukos S., Ward T., Zhu W.-J. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics – ACL '02*, 2001, p. 311. <https://doi.org/10.3115/1073083.1073135>.
- 6 Xu, Wei, Napoles, Courtney, Pavlick, Ellie, Chen, Quanze, Callison-Burch, Chris. *Transactions of the Association for Computational Linguistics*, 2016, pp. 401–415. https://doi.org/10.1162/tacl_a_00107.
- 7 Narayan S., Gardent C. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014, vol. 1, pp. 435–445. <https://doi.org/10.3115/v1/P14-1041>.
- 8 Sulem E., Abend O., Rappoport A. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 738–744. <https://doi.org/10.18653/v1/D18-1081>.
- 9 Sulem E., Abend O., Rappoport A. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018, vol. 1, pp. 685–696. <https://doi.org/10.48550/arXiv.1810.05022>.
- 10 Qonaqjailyqqa negizdelgen qazaq ultiq ashanasy, soyle.kz, 2024. Available: <https://www.soyle.kz/article/view?id=879>. [Accessed: 26-Nov-2024] [in Kazakh]
- 11 Janfada B., Minaei-Bidgoli B. *6th International Conference on Web Research (ICWR)*, 2020, p. 271. <https://doi.org/10.1109/ICWR49608.2020.9122325>.

^{1*}Нұрсана С.Т.,

магистрант, ORCID ID: 0009-0005-7104-4646

*e-mail: samal200174@gmail.com

¹Уалиева И.М.,

ф.-м.ғ.к., қауымдастырылған профессор, ORCID ID: 0000-0003-3853-8896,

e-mail: ualiyeva.irina@kaznu.kz

¹әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қ., Қазақстан

BLEU ЖӘНЕ SARI МЕТРИКАЛАРЫН ҚАЗАҚ ТІЛІНДЕГІ ЖЕҢІЛДЕТІЛГЕН МӘТІНДЕРДІ БАҒАЛАУДА ҚОЛДАНУ: ТАЛДАУ ЖӘНЕ ТИІМДІЛІК

Андатпа

Бұл мақалада қазақ тіліндегі жеңілдетілген мәтіндердің сапасын бағалау әдістемесі BLEU және SARI метрикаларын қолдану арқылы қарастырылады. Мәтіндерді жеңілдету – ақпараттың қолжетімділігін қамтамасыз ету және қазақ тілінде оқу процесін жеңілдету үшін маңызды аспект. BLEU метрикасы,

аударма мен эталонды n-граммаларды салыстыруға негізделген, машиналық аударма сапасын бағалау үшін кеңінен қолданылады, бірақ ол кіріс мәтінінің контекстін ескермейді. SARI метрикасы, мәтіндерді жеңілдетуді бағалауға арнайы жасалған, семантикалық өзгерістерді ескереді және адам бағаларымен жоғары корреляцияны көрсетеді. Зерттеу аясында күрделі сөздерді қарапайым синонимдермен алмастыру және күрделі фразаларды ауыстыру немесе жою алгоритмдері қолданылды. Талдау нәтижелері SARI метрикасы жеңілдетілген мәтіндерде енгізілген өзгерістерге BLEU-ге қарағанда сезімтал екенін көрсетті. Осылайша, BLEU және SARI метрикаларын біріктіріп қолдану қазақ тіліндегі жеңілдетілген мәтіндердің сапасын жан-жақты және дәл бағалауға мүмкіндік береді.

Тірек сөздер: мәтіндерді жеңілдету, қазақ тілі, бағалау метрикалары, BLEU, SARI, мәтін сапасы, машиналық аударма, мәтінді автоматты түрде жеңілдету, жеңілдету алгоритмдері.

¹*Nursapa S.T.,

Master's student, ORCID ID: 0009-0005-7104-4646

*e-mail: samal200174@gmail.com

¹Ualiyeva I.M.,

Cand. Phys.-Math. Sc., Associate Professor, ORCID ID: 0000-0003-3853-8896,

e-mail: ualiyeva.irina@kaznu.kz

¹Al-Farabi Kazakh National University, Almaty, Kazakhstan

APPLICATION OF BLEU AND SARI METRICS IN EVALUATING SIMPLIFIED TEXTS IN KAZAKH: ANALYSIS AND EFFECTIVENESS

Abstract

This article explores the methodology for evaluating the quality of simplified texts in Kazakh using BLEU and SARI metrics. Text simplification is an important aspect for ensuring information accessibility and facilitating the learning process in Kazakh language. The BLEU metric, based on comparing n-grams of translation and reference, is widely used for evaluating the quality of machine translation, but it does not take the context of the input text into account. The SARI metric, specifically designed for evaluating text simplification, considers semantic changes and shows a higher correlation with human judgments. In this study, algorithms for replacing complex words with simple synonyms and for replacing or removing complex phrases were applied. The analysis results showed that the SARI metric is more sensitive to the changes made in simplified texts compared to BLEU. Therefore, the combined use of BLEU and SARI metrics provides a comprehensive and accurate evaluation of the quality of simplified texts in Kazakh.

Key words: text simplification, Kazakh language, evaluation metrics, BLEU, SARI, text quality, machine translation, automatic text simplification, simplification algorithms

Дата поступления статьи в редакцию: 30.12.2024