

O‘ZBEKISTON RESPUBLIKASI RAQAMLI
TEKNOLOGIYALAR VAZIRLIGI OLIY TA‘LIM, FAN VA
INNOVATSIYALAR VAZIRLIGI
MIRZO ULUG‘BEK NOMIDAGI O‘ZBEKISTON MILLIY
UNIVERSITETI



RAQAMLI TEKNOLOGIYALAR VA
SUN'Y INTELLEKTNI RIVOJLANTIRISH
ILMIY-TADQIQOT INSTITUTI



**«KOMPYUTER LINGVISTIKASINING
ZAMONAVIY TEKNOLOGIYALARI – CTCL .2023»**

VAZIRLIK MIQYOSIDAGI ILMIY-NAZARIY ANJUMANI
ILMIY MAQOLALARI VA TEZISLARI
TO‘PLAMI
(2023-YIL 14-APREL)



TOSHKENT – 2023

KOMPYUTER LINGVISTIKASINING ZAMONAVIY TEXNOLOGIYALARI.
Vazirlik miqyosidagi ilmiy-nazariy anjumani ilmiy maqolalari va tezislari to‘plami (2023-yil 14-aprel). –T.: UzMU, 2023. – 288 bet.

Mazkur ilmiy-amaliy anjuman maqolalari va tezislari to‘plamida kompyuter lingvistikasi korpus lingvistikasining dolzarb masalalari, til ta’limida kompyuter texnologiyalari, NLPda til va nutq tahlili (Lingvistik analizator: morfologik, sintaktik va semantik tahlil, nutq analizi va sintezi), mashina tarjimasining lingvistik va dasturiy ta’minoti, kompyuter leksikografiyasi: ontologiya, tezaurus kabi masalalarga bag‘ishlangan tezislardan tashkil topgan.

Ushbu ilmiy maqola va tezislari to‘plami kompyuter lingvistikasi, kompyuter texnologiyalari, amaliy filologiya kabi yo‘nalishida ilmiy izlanishlar olib borayotgan professor-o‘qituvchilar, doktorantlar, mustaqil izlanuvchilar, magistr va talabalar uchun mo‘ljallangan.

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ.
Сборник научных статей и тезисов научно-теоретической конференции (14 апреля 2023 г.). – Т.: НУУ, 2023. – 288 с.

В сборник материалов научно-прикладной конференции вошли статьи и тезисы по актуальным дисциплинам компьютерной лингвистики, а именно: корпусной лингвистике, компьютерным технологиям в языковом образовании, анализу языка и речи в НЛП (Лингвистический анализатор: морфологический, синтаксический и семантический анализ, анализ и синтез речи), лингвистика и лингвистика программ машинного перевода, компьютерная лексикография: онтология, тезаурус.

Материалы научно-прикладной конференции предназначены для профессоров, преподавателей, докторантов, независимых исследователей, магистрантов и бакалавров, работающих в области компьютерной лингвистики, компьютерных технологий, прикладной филологии.

Mas’ul muharrir:

f.f.d., prof.v.b. N. Abduraxmonova.

Ответственный редактор:

д.ф.н., и.о.проф. Н.Абдурахмонова

Tahrir hay’ati:

F.f.n., prof. Z. Tahirov;

F.f.n., dots. D. Fattaxova;

Katta o‘qit. N. Inogamova;

Katta o‘qit. N. Bekmuxamedova;

O‘qit. F. Topildiyeva;

O‘qit. O.Yulbarsov;

O‘qit. F.Aliyev;

O‘qit. M.Jo‘raqulova.

Редакционная коллегия:

к.ф.н., проф. Тахиров З.;

к.ф.н., доц. Фаттахова Д.;

препод. Инагомова Н.;

препод. Бекмухамедова Н.;

препод. Топилдиева Ф.;

препод. Юлбарсов О.;

препод. Алиев Ф.;

препод. Журакулова М.

To‘plamga kiritilgan maqolalar va tezislari mazmuni hamda sifatiga mualliflar mas’uldir.

Авторы несут ответственность за содержание и качество статей и тезисов, включенных в сборник.

MUNDARIJA

1-SHO‘BA

KORPUS LINGVISTIKASINING DOLZARB MASALALARI

N.Z. ABDURAXMONOVA, N.B. TULYAGANOVA. TA‘LIMY KORPUS YARATISH TAJRIBASIDAN	7
S.A. KARIMOV, M.S. TURSUNOV. UZBEKCORPORA.UZ: O‘ZBEK TILI MILLIY KORPUSIDA MATN BILAN ISHLASH	13
Г.И. ТОИРОВА. СОЗДАНИЕ И ЗНАЧЕНИЕ ЯЗЫКА КОРПУС В УЗБЕКИСТАНЕ.....	22
N.Z. ABDURAXMONOVA, K. RAJABOVA, S. XASANOV, D. JUMANAZAROVA, D. O‘ROQOVA, R. YO‘LDOSHEVA. OG‘ZAKI NUTQ KOMPYUTER KORPUSLARI.....	27
M.S. SHARIPOV, L.U. KURBANOVA, R.U. QURBANOVA. O‘ZBEK TILI KORPUSI DASTURIY TA‘MINOTINI YARATISH	33
M.A. ABJALOVA. PARALLEL KORPUSLARNI YARATISH ASOSLARI	37
G.X. NURMATOVA. DISCUSSION OF NANO-PREFIXED TERMS	44
N. TOSHPO‘LATOVA. KORPUS LINGVISTIKASINING PROGRESSIV MASALALARI	48
Г.Н. ХИДИРОВА. О НЕОБХОДИМОСТИ СОЗДАНИЯ НАЦИОНАЛЬНОГО КОРПУСА УЗБЕКСКОГО ЯЗЫКА.....	52
K.Q. VASLIDDINOVA. KORPUS LINGVISTIKASI VA DUNYO TILSHUNOSLIGIDA KORPUSGA OID NAZARIYA VA TADQIQOTLAR.....	57
O.O. YULBARSOV, Z.H. RUSTAMOVA. KORPUS–TABIY TILLARNING MILLIY XUSUSIYATINI SAQLAB QOLISHDA MUHIM INSTRUMENT	61
Z. ISMADIYOROV, T. KAMILOV. SEMANTIK ANNOTATSIYALANGAN KORPUSLAR TAVSIFI	65
M.K. SADIKOVA. KOMMUNIKATIV TIL IMKONIYATLARINI RIVOJLANTIRISHDA KORPUS TEXNOLOGIYALARIDAN FOYDALANISH	68
G. ARABBOYEVA. USING DIGITAL METADATA IN ELECTRONIC CORPORA	72
I.I. ANORBOYEVA, M. JO‘RAQULOVA. PARALLEL KORPUSNING O‘RGANILISHIGA DOIR TAJRIBALAR TAHLILI.....	77
D.K. TOSHPO‘LATOVA. ILMIY USLUBDA SOHAVIY TERMINOLOGIYANI YARATISHDA VALENTLIK TUSHUNCHASI	82
S.I. QO‘ZIYEVA. O‘ZBEK KORPUS LINGVISTIKASIDA XI-XIX ASR ADABIY MANBALARI LEKSIKOGRAFIYASINING O‘RNI.....	89
K. MUKHAMMADIYEVA, M. MUKHAMMADJONOVA, D. RAVSHANOVA, N. SHARIPOVA. UZBEK ELECTRONIC CORPUS OF NEWSPAPER TEXTS.....	92
T.R. YANDASHOVA. INGLIZ VA O‘ZBEK TILI FRAZEMALARIDA INSON QIYOFASI VA XULQ-ATVORI IFODASI.....	96

A.SH. ISMAILOV, G.T. QODIROVA, M.A. YIGITALIYEVA, R. XUSANOVA. O‘ZBEK STEMASI ALGORITMI	101
M.X. ESHMAMATOVA. KORPUS LINGVISTIK TA‘MINOTINI YARATISH MASALALARI	106

2-SHO‘BA

TIL TA‘LIMIDA KOMPYUTER TEXNOLOGIYALARI

G.S. KELDIYOROVA. O‘ZGA TILLARNI O‘QITISHDA INTERNETNING O‘RNI	110
A. TESHABOYEV, M.Z. SAYFUTDINOVA. CHET TILLARINI O‘QITISHDA AXBOROT TEXNOLOGIYALARINING O‘RNI	114
N. BEKMUXAMEDOVA. O‘ZGA TILLI GURUHLARDA TIL O‘QITISHDA MULTIMEDIA TAQDIMOTIDAN FOYDALANISH	118
F.R. TOPILDIYEVA. TA‘LIM JARAYONIDA KOMPYUTER TEXNOLOGIYALARI, MULTIMEDIALI O‘QUV KONTENTLARIDAN FOYDALANISH O‘RNI VA SAMARASI.....	121
D.A. FATTAHOVA. TILLARNI O‘QITISH JARAYONIDA MULTIMEDIALI O‘QUV KONTENTLARIDAN FOYDALANISH	125
Б.Х. ДАНИЯРОВ. АКАДЕМИК ЁЗУВ ДАРСЛАРИДА ЗАМОНАВИЙ СМАРТФОНЛАРДАН САМАРАЛИ ФОЙДАЛАНИШ	135
X.X. AXMEDOVA, D.A. FATTAHOVA. ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННО КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ ПРИ ОБУЧЕНИИ ЯЗЫКУ	138
H.Ш. ИНОГАМОВА. ИСПОЛЬЗОВАНИЕ СОВРЕМЕННЫХ ИКТ В ПРОЦЕССЕ ОБУЧЕНИЯ ЯЗЫКАМ	143
S.B. BOYSARIYEVA. TIL O‘RGANISHDA KYTO‘ VA MYTO‘DAN FOYDALANISH	147
Z.SH. ABDULLAYEVA, G.A.SHIHNAZAROVA. INTELLEKTUAL TAHLIL ASOSIDA CHET TILI BILIMINI TEKSHIRISHNI MODELLASHTIRISH	150

3-SHO‘BA

NLPDA TIL VA NUTQ TAHLILI (LINGVISTIK ANALIZATOR: MORFOLOGIK, SINTAKTIK VA SEMANTIK TAHLIL, NUTQ ANALIZI VA SINTEZI)

E. NAZIROVA, SH. ABIDOVA, M. UZAKOVA. TURKIY TILLAR O‘RTASIDA TABIIY TILLARNI QAYTA ISHLASH MODEL (O‘ZBEK TILI MISOLIDA)	153
Л. БОБОЕВ. МАТНЛИ ХУЖЖАТЛАРНИ ТАСНИФЛАШНИНГ ГИБРИД АЛГОРИТМИНИ ИШЛАБ ЧИҚИШ	159
H. РАВШАНОВ, И.И. БАКАЕВ. ПРОЕКТИРОВАНИЕ MORFOЛОГИЧЕСКОГО АНАЛИЗАТОРА ДЛЯ ЕСТЕСТВЕННЫХ ЯЗЫКОВ НА ПРИМЕРЕ УЗБЕКСКОГО ЯЗЫКА.....	163
U. TULIYEV. ANALYSIS OF THE SEMANTIC RELATEDNESS OF WORDS BASED ON A COMMON DICTIONARY	172
M.YA. URAZALIYEVA. KORPUSDA MORFOLOGIK ANALIZATOR UCHUN LINGVISTIK MODELLASHTIRISH MASALALARI.....	176
A. SHAYMARDANOV. RECENT ADVANCEMENTS IN UZBEK NLP TOOLS.....	181

ПРОЕКТИРОВАНИЕ МОРФОЛОГИЧЕСКОГО АНАЛИЗАТОРА ДЛЯ ЕСТЕСТВЕННЫХ ЯЗЫКОВ НА ПРИМЕРЕ УЗБЕКСКОГО ЯЗЫКА

Равшанов Н., Бакаев И. И.

Научно-исследовательский институт развития цифровых технологий и искусственного интеллекта
 Бухарский государственный университет

Abstract. A brief review of existing morphological analyzers for agglutinative languages is given. An architecture of a morphological analyzer for the Uzbek language is proposed, an object-oriented model and a diagram of which are described using UML.

Keywords. Computational linguistics, UML, tokenization, stemming, morphonology, state machine.

Аннотация. Приведен краткий обзор существующих морфологических анализаторов для агглютинативных языков. Предложен архитектура морфологического анализатора для узбекского языка, объектно-ориентированная модель и диаграмма вариант использования которого описана при помощи UML.

Ключевые слова. Компьютерная лингвистика, UML, токенизация, стемминг, конечный автомат.

Введение.

Морфологический анализатор является неотъемлемой частью решения подзадач поисковых систем (машинного перевода, информационного поиска, извлечения информации), электронных библиотек, архивов, систем «вопрос-ответ» и документооборота. Морфологический анализатор – это программное обеспечение, которое решает ряд комплексных задач, таких как определение грамматических признаков, основ, лемм, морфем и части речи словоформы рис.1.



Рис.1. Формальная схема морфологического анализатора

Для создания морфологического анализатора существует множество подходов: конечный автомат, конечный автомат с выходом, детерминированный ациклический конечный автомат, подходы основанные на правилах, корпусах, шаблонах, двухуровневой

морфологии и гибридных подходов. На основе их создано множество готовых решений для групп флективных и агглютинативных языков мира Рис.2.

Только определенное количество этих естественных языков обеспечено необходимыми лингвистическими ресурсами и инструментами морфологического анализатора для обработки естественного языка.

Большинство научных исследований основано на комбинаторной смеси нескольких морфологических подходов. Так как в статье будем проектировать морфологический анализатор на примере узбекского языка, поэтому ниже описываются некоторые исследования, относящиеся к агглютинативным языкам.

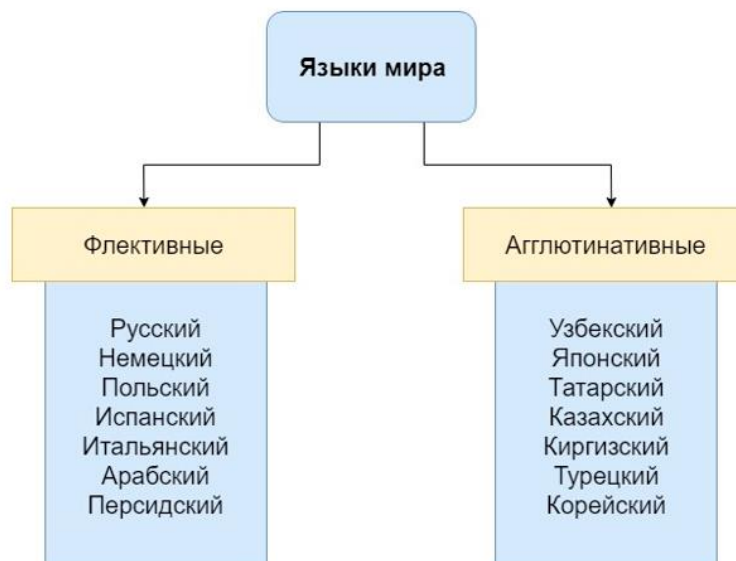


Рис.2. Перечень флективных и агглютинативных языков мира

Связанные работы

Выполняется ряд научных работ по таким языкам, как турецкий, казахский, туркменский, узбекский, киргизский, татарский, относящимся к семье тюркских языков, принадлежащим к группе агглютинативных языков.

В научных работах по турецкому языку были разработаны такие морфологические анализаторы, как TRmorph[1] и TRMOR [2], основанные на подходе двухуровневой морфологии, которые реализованы с помощью SFST. TRmorph состоит трёх компонентов: конечного автомата, представленного регулярными выражениями, набора правил фонологии и орфографии, а также словаря из 37101 слов, принадлежащих 9-ти категориям частей речи. Лексикон морфологического анализатора TRMOR составлен из ресурса Википедии. Когда TRMOR и TRmorph были протестированы с использованием 108 слов, TRMOR дал точность 72%, а TRmorph дал точность 38%[2].

Также авторами [3] описан морфологический анализатор двухуровневой морфологии для турецкого языка, который состоит из пяти модулей: конечный автомат с выходом, механизм правил для суффиксов (файл в формате .xml), лексикон (база данных), структура данных виде дерева и кэш типа LRU (least recently used). Лексикон состоит 54 000 слов, из них 19 000 относится к именам существительных.

В исследованиях [4], [5] разработаны морфологические анализаторы на основе правил для казахского языка. Для реализации правил чередования и морфотактики двухуровневый морфологии используется конечный автомат с выходом Foma и XFST (Xerox Finite State Tools). Основная функция морфологического анализатора устранение неоднозначности для казахского языка, который является разновидностью теггера Брилла.

Также для туркменского языка [6] разработан морфологический анализатор с подходом двухуровневой морфологии. Грамматические правила языка реализованы с помощью конечного автомата типа XFST.

Авторами изложено [7] проектирование и разработка комплекса программных средств который включает лингвистические ресурсы (тезаурусы и онтологии) и инструменты обработки текстов (генератор словоформ, морфологический анализатор и средство морфологического устранения неоднозначности слов) на основе корпусного подхода для казахского языка.

В статье [8] описывается разработанный морфологический анализатор для казахского, татарского и кумыкского языков. Для реализации морфотактических правил используется конечный автомат типа HFST (Helsinki Finite-State Toolkit). Общее количество слов в лексиконе для казахского языка составляет 11224, для татарского – 10737, для кумыкского 4845 слов.

В работе [9] обсуждается инструмент морфологического анализатора для узбекского языка под названием MorphUz. MorphUz работа на основе подхода двухуровневой морфологии, данный подход может определить основу слов [10] и анализировать суффиксы.

В исследовании [11] разработан инструмент морфологического анализа UZMORPP для узбекского языка. Инструмент включает морфотактические и морфофонемные правила узбекского языка, который реализован с помощью логического программирования на языке программирования Prolog.

В работе [12] предложен электронный словарь окончаний узбекских слов для определения морфологических основ слов. Электронный словарь проанализирован с помощью казахского морфологического анализатора типа стемминг [13], [14].

Хотя в представленных выше работах были получены важные результаты фундаментального и практического характера, они не были спроектированы как полноценная система морфологического анализа. Поэтому цель этой задачи спроектировать системы для всех задач морфологического анализа узбекского языка.

Теоретический анализ

Словарный состав узбекского языка постоянно пополняется новыми словами. Пополнение происходит путём образования новых слов с помощью аффиксации и композиции. Аффиксация и композиция являются основными способами образования слов. Аффиксальный способ образования слов наиболее продуктивный в узбекском языке. В этом способе новые слова создаются путём присоединения к корню слова образующих аффиксов. Композиционный способ — это создание новых слов путем сложения двух и более слов. В результате двух указанных выше способов в узбекском языке по структуре и формированию образуются следующие типы слов Рис.3.



Рис.3. Типы слов в узбекском языке по их структуре.

В современном узбекском языке есть шесть самостоятельных частей речи: Имя существительное, Имя прилагательное, Имя числительное, местоимение, глагол, наречие. Каждая часть речи по структуре имеет вышеуказанные типы слов.

На основании вышеприведенной теоретической информации мы определили следующие сущности:

- слова;
- типы слов;
- части речи;
- аффикс;
- типы аффиксов.

Далее определяем сущности для задач морфологического анализатора. К ним относятся следующие:

- Разбиение текста на токены;
- Определение основы слов;
- Определение морфем и грамматических признаков.

Постановка задачи

С учетом проведенного анализа в статье решаются следующие задачи:

- Разработка модели классов морфологического анализатора;
- Построение диаграммы варианта использования для определённых групп пользователей;

Метод решение

Исходя из выше приведенных теоретических материал определяем следующие классы рис.4:

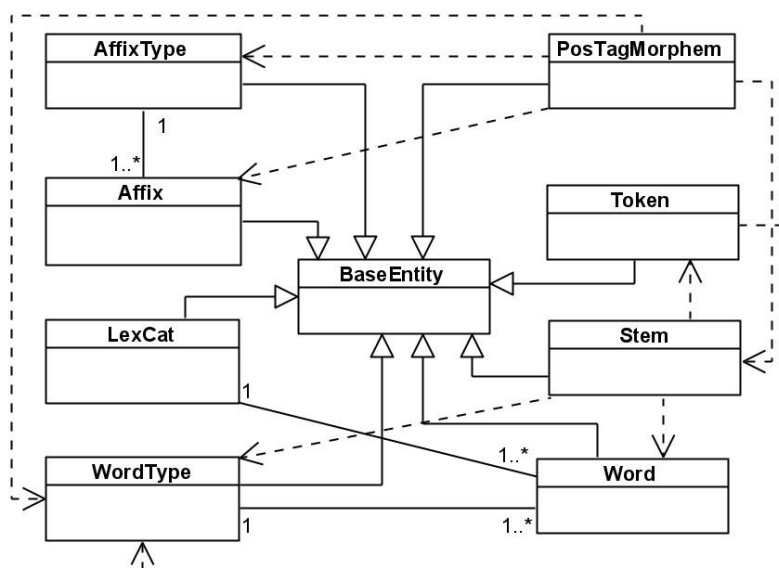


Рис.4. Объектно-ориентированная модель морфологического анализатора

«BaseEntity» — это абстрактный класс, все классы наследуются от этого класса. Этот класс состоит из полей, представляющих идентификатор(id) и значение(value).

«Words» — это класс, представляющий сущность «Слова». У класса есть свои поля «WordParent», чтобы определить, из какого слова оно было создано.

«WordType» — это класс, представляющий сущность «Типы слов».

«Affix» — это класс, представляющий сущность «Аффиксы».

«AffixType» — это класс, представляющий сущность «Типы аффиксов».

«LexCat» — это класс, представляющий сущность «Лексические категории» то есть части речи.

Все классы представленных в диаграммы классов наследоваться от класса BaseEntity. Для «WordType», «Affix», «AffixType», «LexCat» классов достаточно полей, унаследованных от класса BaseEntity. Поэтому у этих классов нет собственных полей. Эти поля могут использоваться для хранения названия типа слов, аффиксов, типа аффиксов, лексических категорий и их идентификаторов.

«Word_LexCat» — это класс, представляющий взаимосвязь между классами «Words» и «LexCat». Связь между классами «Words» и «LexCat» использует дополнительный класс «Word_LexCat», чтобы избежать омонимии некоторых слов.

Связь между классами Affix и LexCat использует класс Affix_LexCat для определения омонимичных суффиксов.

«Stem» — это класс, представляющий сущность «основы слов». Класс состоит из следующих полей и методов:

- "OrfoTypeWord" - метод определения типа слова;
- «StemmingWordW2» - метод, находящий основу слова (основу) типа w_{sb} ;
- "StemmingWordW3W4" - метод нахождения основы (основы) слов типа w_j и w_t ;
- метод "StemmingWordW5", находящий основу слова (основу) типа w_{ab} ;

- метод "StemmingWord", находящий основу (основу) слов типа w_{st} , w_{sy} и w_{psy} ;
- «Lemmatization» - метод, который преобразует словоформы в лемму типа w_{st} , w_{sy} и

w_{psy} ;

- "isprefix" - метод, определяющий, есть ли у слова префикс или нет;
- "prefixLen" - метод определения длины префикса;
- "ReadRule" – метод чтения правил;

Алгоритмы StemmingWordW2, StemmingWordW3W4, StemmingWordW5 и StemmingWord методов реализован в работе [15] ранее.

«Token» — это класс, который разбивает текст на слова. Класс содержит метод «Tokenlash»[16] , который разбивает текст на слова.

«POSTagMorphem» — это класс, который определяет слово в морфемы и свойства морфем[17]. Класс состоит из следующих методов:

- «isbelogswordlang» — метод, проверяющий принадлежность слова к языку;
- "POSTaggingMorphemic" - метод маркировки аффиксов грамматическими тегами;
- "MorphemicList" - метод, формирующий список аффиксов;
- "MorphemicTypeList" - метод, формирующий тип аффиксов;

Класс «POSTagMorphem» использует отношение «Dependency» (зависимость) с классами «AffixType», «Affix» и «WordType», то есть в качестве входных и выходных значений.

Далее построим диаграмму вариант использования для определённых групп пользователей.

Для системы морфологического анализатора можно выделить следующие группы пользователей:

- пользователи (студент, преподаватель, API пользователь) – может провести морфологический анализ (задачи токенизация, определение основ слов, морфем и грамматических признаков) слов с помощью системы рис.5а.

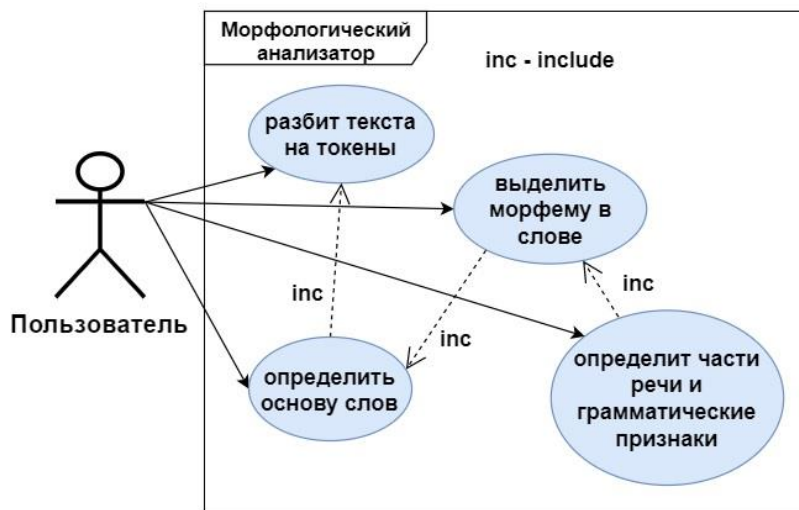


Рис.5а. Диаграмма вариант использования морфологического анализатора для пользователей

- эксперт-лингвист – специалист в области компьютерный лингвистики который может сделать операция CRUD (аббревиатура англ. слов create, read, update, delete) рис.5b.

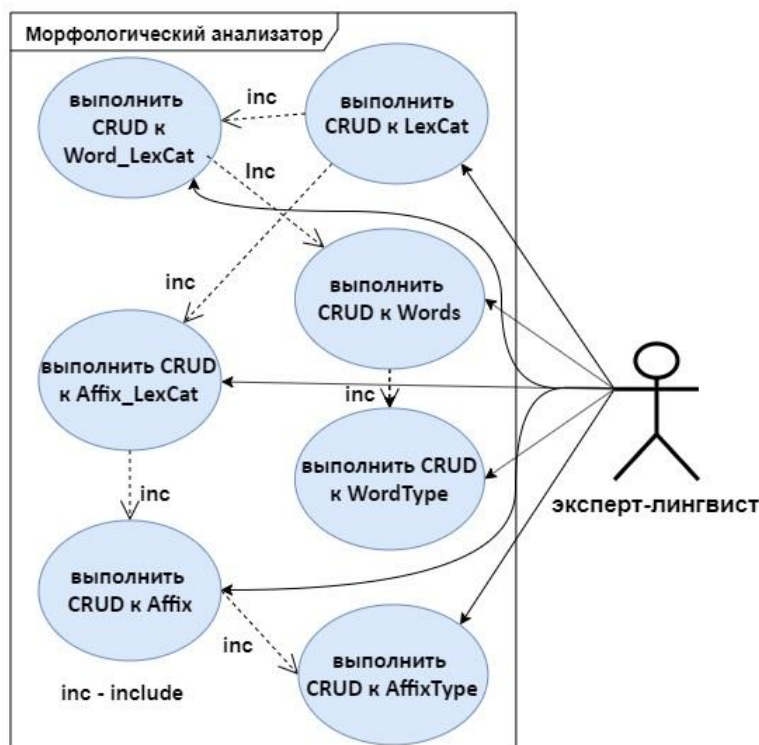


Рис.5b. Диаграмма вариант использования морфологического анализатора для эксперт-лингвистов.

Результаты

На основе применения разработанного морфологического анализа слов узбекского языка для текстовых запросов на узбекском языке при поиске информационных ресурсов в системе электронной библиотеки Бухарского областного информационно-библиотечного центра имени Абу Али ибн Сино и в центре информационных ресурсов Бухарского государственного университета обеспечена возможность сокращения времени и трудозатрат на 9-11% при выполнении операций составления библиографических описаний документов на узбекском языке в системе электронного каталога за счет автоматической коррекции орфографических ошибок и выдачи правильного варианта написания слов.

Заключение

По итогам проведённого анализа научных исследований, изложенных в работе, связанных с созданием морфологических анализатора для узбекского языка, относящихся к семье агглютинативных языков, разработаны модели классов морфологического анализатора, и построена диаграмма варианта использования морфологического анализатора для определённых групп пользователей.

Использованная литература:

1. G. Gagri, “A Freely Available Morphological Analyzer for Turkish,” in *LREC*, 2010, pp. 820–827. Accessed: Jun. 29, 2020. [Online]. Available: <http://www.gnu.org/licenses/gpl.html>
2. K. Ayla, S. Helmut, T. Ahmet E, and K. Özkan, “TRMOR: a finite-state-based morphological analyzer for Turkish,” *Turkish J. Electr. Eng. Comput. Sci.*, vol. 27, pp. 3837–3851, 2019, doi: 10.3906/elk-1902-125.
3. O. T. Yıldız, A. Begüm, and E. Gökhan, “An Open, Extendible, and Fast Turkish Morphological Analyzer,” in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 2019, pp. 1364–1372.
4. G. Kessikbayeva and I. Cicekli, “A Rule Based Morphological Analyzer and A Morphological Disambiguator for Kazakh Language,” *Linguist. Lit. Stud.*, vol. 4, no. 1, pp. 96–104, 2016.
5. G. Kessikbayeva and I. Cicekli, “Rule Based Morphological Analyzer of Kazakh Language,” in *Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM*, 2014, vol. Proceeding, pp. 46–54.
6. A. C. Tantuğ, E. Adalı, and K. Oflazer, “Computer Analysis of the Turkmen Language Morphology,” in *Advances in Natural Language Processing*, 2006, pp. 186–193.
7. D. Akhmed-Zaki, M. Mansurova, G. Madiyeva, N. Kadyrbek, and M. Kyrgyzbayeva, “Development of the information system for the Kazakh language preprocessing,” *Cogent Eng.*, vol. 8, no. 1, p. 1896418, 2021, doi: 10.1080/23311916.2021.1896418.
8. J. Washington, I. Salimzyanov, and F. Tyers, “Finite-state morphological transducers for three Kypchak languages,” in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC’14)*, 2014, pp. 3378–3385. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2014/pdf/1207_Paper.pdf
9. N. Abdurakhmonova, I. Alisher, and R. Sayfulleyeva, “MorphUz: Morphological Analyzer for the Uzbek Language,” in *2022 7th International Conference on Computer Science and Engineering (UBMK)*, 2022, pp. 61–66. doi: 10.1109/UBMK55850.2022.9919579.
10. D. Mengliev, V. Barakhnin, and N. Abdurakhmonova, “Development of Intellectual Web System for Morph Analyzing of Uzbek Words,” *Appl. Sci.*, vol. 11, no. 19, 2021, doi: 10.3390/app11199117.
11. G. Matlatipov and V. Zygmunt, “Representation of Uzbek Morphology in Prolog,” *Asp. Nat. Lang. Process.*, vol. 5070, pp. 83–110, 2009, doi: 10.1007/978-3-642-04735-0.
12. S. Matlatipov, U. Tukeyev, and M. Aripov, “Towards the Uzbek Language Endings as a Language Resource,” in *Advances in Computational Collective Intelligence*, 2020, pp. 729–740.
13. U. Tukeyev, A. Turganbayeva, B. Abduali, D. Rakhimova, D. Amirova, and A. Karibayeva, “Lexicon-free stemming for Kazakh language information retrieval,” in *2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*, 2018, pp. 1–4. doi: 10.1109/ICAICT.2018.8747021.
14. U. Tukeyev, A. Karibayeva, A. Turganbayeva, and D. Amirova, “Universal Programs for Stemming, Segmentation, Morphological Analysis of Turkic Words,” in *Computational Collective Intelligence*, 2021, pp. 643–654.

15. I. . Bakayev, “Development of a stemming algorithm based on a linguistic approach for words of the uzbek language,” in *International Conference on Scientific, Educational & Humanitarian Advancements*, 2021, pp. 195–202.
16. I. I. Bakaev, “Linguistic features tokenization of text corpora of the Uzbek language,” *Bull. TUIT Manag. Commun. Technol.*, vol. 4, 2021
17. Н. Равшанов, И. . Бакаев, and Т. . Шафиев, “Особенности разработки морфологического анализатора узбекского языка,” *Проблемы вычислительной и прикладной математики*, vol. 4, no. 28, pp. 121–131, 2020.