OSIYO TEXNOLOGIYALAR UNIVERSITETI

YASHIL

IQTISODIYOT VA TARAQQIYOT

Ijtimoiy, iqtisodiy, texnologik, ilmiy, ommabop jurnal

№6-MAXSUS SON

OPEN ACCESS

INDEX COPERNICUS INTERNATIONAL

Academic Resource Index ResearchBib

OpenAIRE

ULRICHSWEB GLOBAL SERIALS DIRECTORY

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INTERNATIONAL CENTRE

BASE

ROAD

Crossref

НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА eLIBRARY.RU

74-91 xalqaro daraja

ISSN: 2992-8982

YASHIL
IQTISODIYOT&TARAQQIYOT
Ijtimoiy, iqtisodiy, texnologik, ilmiy, ommabop jurnal

Elektron nashr. 153 sahifa.
E'lon qilishga 2025-yil 1-iyunda ruxsat etildi.

IQTISODIYOT&TARAQQIYOT

Ijtimoiy, iqtisodiy, texnologik, ilmiy, ommabop jurnal

08.00.01 Iqtisodiyot nazariyasi
08.00.02 Makroiqtisodiyot
08.00.03 Sanoat iqtisodiyoti
08.00.04 Qishloq xo‘jaligi iqtisodiyoti
08.00.05 Xizmat ko‘rsatish tarmoqlari iqtisodiyoti
08.00.06 Ekonometrika va statistika
08.00.07 Moliya, pul muomalasi va kredit
08.00.08 Buxgalteriya hisobi, iqtisodiy tahlil va audit
08.00.09 Jahon iqtisodiyoti
08.00.10 Demografiya. Mehnat iqtisodiyoti
08.00.11 Marketing
08.00.12 Mintaqaviy iqtisodiyot
08.00.13 Menejment
08.00.14 Iqtisodiyotda axborot tizimlari va texnologiyalari
08.00.15 Tadbirkorlik va kichik biznes iqtisodiyoti
08.00.16 Raqamli iqtisodiyot va xalqaro raqamli integratsiya
08.00.17 Turizm va mehmonxona faoliyati

Muassis: “Ma’rifat-print-media” MChJ
Hamkorlarimiz: Toshkent davlat iqtisodiyot universiteti, O‘zR Tabiat resurslari vazirligi, O‘zR Bosh prokuraturasi huzuridagi IJQK departamenti.

Jurnalning ilmiyligi:
““Yashil” iqtisodiyot va taraqqiyot” jurnali
O‘zbekiston Respublikasi Oliy ta’lim, fan va innovatsiyalar vazirligi huzuridagi Oliy attestatsiya komissiyasi rayosatining 2023-yil 1-apreldagi 336/3-sonli qarori bilan ro‘yxatdan o‘tkazilgan.

MUNDARIJA

MUNDARIJA СОДЕРЖАНИЕ CONTENTS

MUNDARIJA СОДЕРЖАНИЕ CONTENTS

ALGORITHMS AND SOFTWARE FOR AUTOMATIC SPELLING AND GRAMMATICAL EDITING OF UZBEK WORDS

Daminov Sunatullo Furqat ugli
Asian University of Technologies
Master's student in the field of Information Technologies in Education
E-mail: sdf6971@gmail.com

Eshkarayeva Narkhol G'uzarovna
Karshi State University, PhD, associate professor
E-mail: 1970narxol@gmail.com

Boymurodov Farrukh Farkhod ugli
Karshi State University, PhD student
E-mail: farruxboymurodov96@gmail.com

Abstract: In the process of globalization and digital informatization, information technologies play a very important role in the development of the Uzbek language. As a result of the Natural Language Processing (NLP) process of computer linguistics and the creation of a (formal) form of natural language that can be understood by a computer, language-related issues (editing, analysis, translation, voicing electronic text, converting spoken speech into electronic text, communicating with a robot , such as converting large text into compact text) are being created. This article reviews the UZMORF program and tools written in the Python programming language for natural language processing (NLP), which is one of the main areas of computational linguistics. This program was created to automate the editing and processing of Uzbek words and their spelling and grammar. The system works on texts and performs syntactic and morphological analysis on words. As a result of these analyses, the system determines the words and their grammatical solutions and marks the places where there are errors.

Key words: NLP, orthographic correction, grammatical correction, uzmorf.

Annotatsiya: Globallashuv va raqamli axborotlashtirish jarayonida o'zbek tilining rivojlanishida axborot texnologiyalari juda muhim o'rin tutadi. Kompyuter tilshunosligining tabiiy tilni qayta ishlash (NLP) jarayoni va kompyuter tomonidan tushuniladigan tabiiy tilning (rasmiy) shaklini yaratish natijasida til bilan bog'liq muammolar (tahrirlash, tahlil qilish, tarjima qilish, elektron matnni ovozli qilish, og'zaki nutqni elektron matnga aylantirish, robot bilan muloqot qilish, masalan, katta matnni ixcham matnga aylantirish) yaratiladi. Ushbu maqolada hisoblash tilshunosligining asosiy yo'nalishlaridan biri bo'lgan tabiiy tilni qayta ishlash (NLP) uchun Python dasturlash tilida yozilgan UZMORF dasturi va vositalari ko'rib chiqiladi. Bu dastur o'zbek tilidagi so'zlarni hamda ularning imlo va grammatikasini tahrirlash va qayta ishlashni avtomatlashtirish maqsadida yaratilgan. Tizim matnlar ustida ishlaydi va so'zlar ustida sintaktik va morfologik tahlil qiladi. Ana shunday tahlillar natijasida tizim so'zlar va ularning grammatik yechimlarini aniqlab, xatosi bor joylarni belgilab beradi.

Kalit so'zlar: NLP, orfografik tuzatish, grammatik tuzatish, uzmorf.

Аннотация: В процессе глобализации и цифровой информатизации информационные технологии играют очень важную роль в развитии узбекского языка. В результате процесса обработки естественного языка (NLP) компьютерной лингвистики и создания (формальной) формы естественного языка, которая может быть понята компьютером, создаются языковые вопросы (редактирование, анализ, перевод, озвучивание электронного текста, преобразование устной речи в электронный текст, общение с роботом, например, преобразование большого текста в компактный текст). В данной статье рассматривается программа UZMORF и инструменты, написанные на языке программирования Python для обработки естественного языка (NLP), который является одним из основных направлений компьютерной лингвистики. Эта программа была создана для автоматизации редактирования и обработки узбекских слов, их орфографии и грамматики. Система работает с текстами и выполняет синтаксический и морфологический анализ слов. В результате этих анализов система определяет слова и их грамматические решения и отмечает места, где есть ошибки.

Ключевые слова: НЛП, орфографическая коррекция, грамматическая коррекция, узморф.

In order for natural language text to be of high quality, text processing needs to be fully analyzed. However, the difficulties that arise in creating such an analysis are that in practice all the theoretical rules developed so far have not yet been implemented. The main problems here are the complexity of text analysis and the complexity of creating a system that implements a complete model of the world's languages. A text analysis system should be able to analyze the text entered by the user in terms of syntax (sentence structure), semantics (concepts used in the text) and pragmatics (correct use of concepts) [1]. Our system has the ability to detect and correct spelling errors in the text entered by the user. In this case, the system must create a suitable internal representation and synthesize its answer in natural language to infer the result. In general, the system that supports text analysis in this article, the UZMORF program, includes the modules "Spelling Analysis" and "Grammatic Editing".

The UZMORF program is a program created for automatic processing of texts in the Uzbek language. This program analyzes Uzbek words and their morphological structure and helps to prepare it for processing on computers. The operating principles of the UZMORF program are as follows:

Words and their morphological structure: UZMORF determines the morphological structure of each word in the Uzbek language. This structure information includes information related to which part of the word, which gender it belongs to, and their form.

Morphological analysis: The program is based on language rules and features prepared on the morphology of the Uzbek language. Uses morphological analysis methods to analyze words and determine their morphological information.

Utilities and modules: UZMORF is written in the Python programming language and uses modules to analyze texts in Uzbek by word and structure. These modules help you work with texts and identify morphological information in them.

Expert Uzbek language database: UZMORF uses a database containing Uzbek words and their morphological information to operate on the texts it is analyzing.

Technical developments: UZMORF software is updated to support new information on new Uzbek words and their morphological structure, taking into account technical developments.

The UZMORF program is used to automatically process texts in the Uzbek language and analyze their morphological structure. This program can be used for editing Uzbek texts, automatic translation, language teaching and other technical tools.

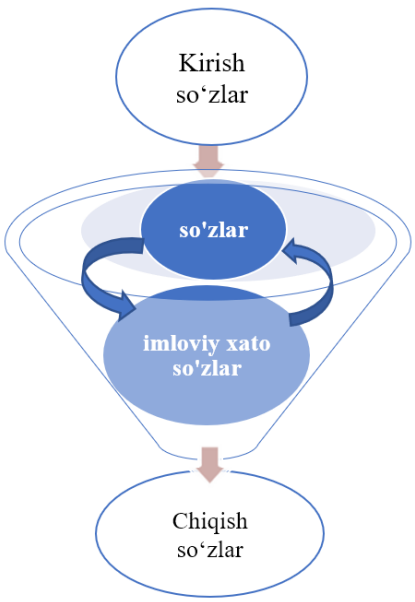
There are several ways to edit Uzbek words grammatically:

Orthography: When editing words, the orthography rules of the Uzbek language should be taken into account. This includes rules such as correct placement of letters, correct use of extra characters and vowels, capitalization of words, and more.

Morphology: Provides editing for morphology, word form, and concepts. This includes editing by word form (singular or plural, person and number, etc.).

Lexicology: Lexicology also plays an important role in the editing process. This provides editing for words and their meanings, their content, and the practical and aesthetic correctness of words or phrases.

These methods help to follow the rules of the Uzbek language in the general editing process.



The process of automatic spelling correction is carried out according to the following algorithm:

1. When you enter words, the system will start searching for the word in the main dictionary. If the word is found in the dictionary, it goes to step 4, otherwise it continues to step

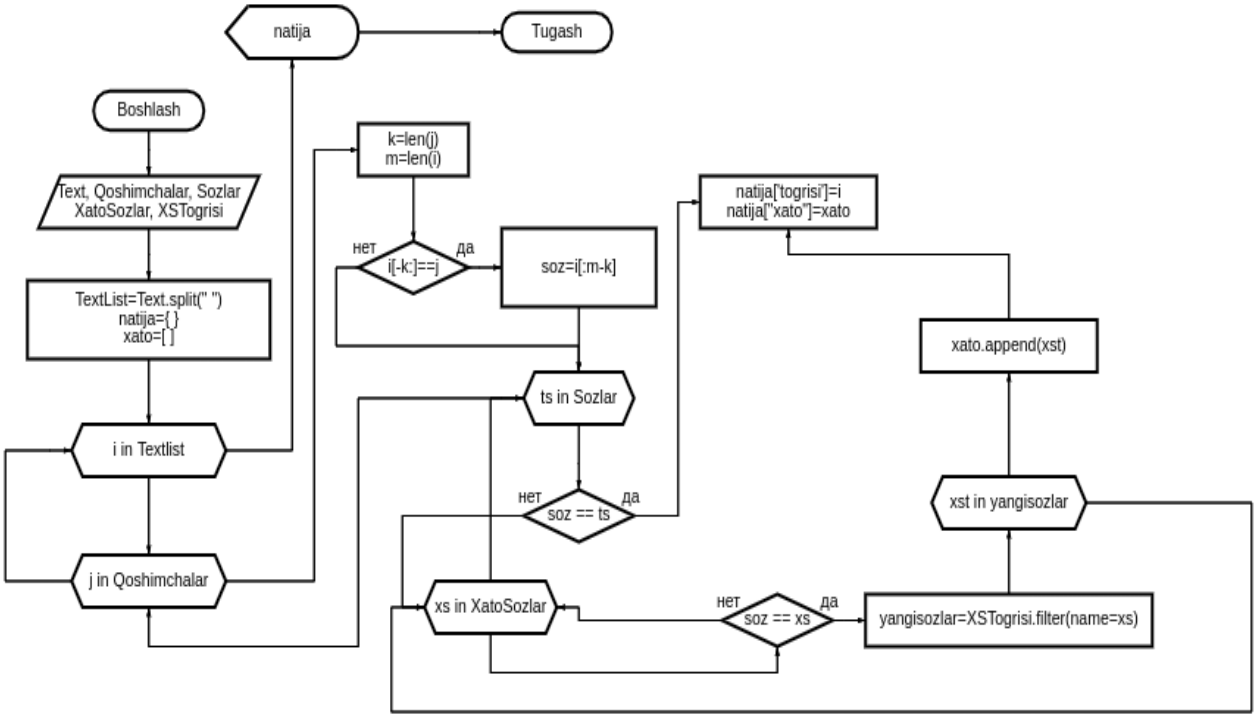
2. In this step, a word is searched from the list of misspelled words, and as a result, the searched word may be more than one. The correct word corresponding to these words is determined, and if the result is successful, it goes to step 1.

3. At this stage, a list of words corresponding to each word in the text found as a result of the search is formed and presented to the user, and the user is transferred to the 4th stage.

4.The user edits the text by selecting the words found by the system as errors from the list of words provided by the system and completes the work of the program.

About 86,000 Uzbek words in Latin script are included in the system database. The misspelled words related to each of these words are attached, that is, the probability of misspelling of each word is at least equal

to the number of letters in it, and all variants are included. A spelling error in the text entered by the user can have several spelling correct forms of the word. The user is able to identify the word that matches the text.



1. Block diagram of UZMORF text editing algorithms.

The user enters the text to be checked and the system edits the text in the following steps:

• the entered text is divided into words and suffixes are extracted from the word;

• each word is checked for availability in the list of word stems available in the database. If it exists, this word is left as it is, otherwise, the misspelled words corresponding to each word are checked from the list of words and the correct variants of the misspelled words are determined. presented to the user in the form of a list;

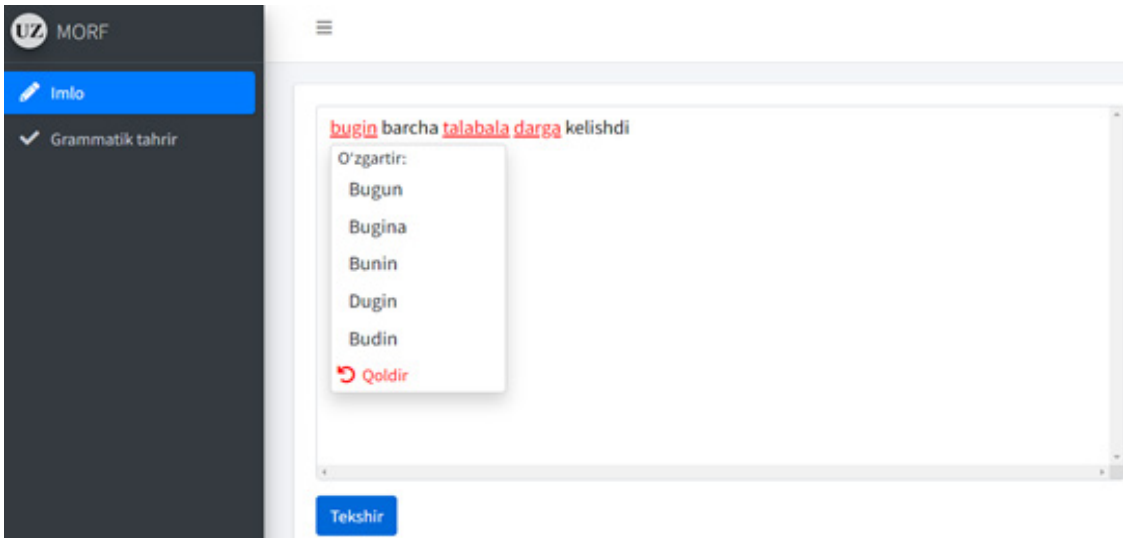
• then the user is allocated words with spelling errors and the user is required to correct the words with errors, i.e. select and change the words that match the entered text.

Artificial intelligence algorithm “Spell Checker” (text checker) algorithms were used in the program to identify errors of words that do not exist in the word base or are no longer used. These algorithms help to identify and correct word errors in texts. That is, it checks each word according to existing dictionaries and grammar rules. If the entered word is not found in the available dictionaries or there are alternatives to edit it, the word checker algorithm will suggest these alternatives. To eliminate these situations, the artificial intelligence algorithm is formulated as follows:

```
import re
from collections import Counter
def words(text):
    return re.findall(r'\w+', text.lower())
def train(features):
    model = Counter()
    for word in features:
        model[word] += 1
    return model
def edits1(word):
    letters    = 'abcdefghijklmnopqrstuvwxyz'
    splits     = [(word[:i], word[i:])    for i in range(len(word) + 1)]
    deletes    = [L + R[1:]               for L, R in splits if R]
    transposes = [L + R[1] + R[0] + R[2:] for L, R in splits if len(R)>1]
```

```
    replaces   = [L + c + R[1:]          for L, R in splits if R for c in letters]
    inserts    = [L + c + R              for L, R in splits for c in letters]
    return set(deletes + transposes + replaces + inserts)
def known(words, model):
    return set(w for w in words if w in model)
def known_edits2(word, model):
    return set(e2 for e1 in edits1(word) for e2 in edits1(e1) if e2 in model)
def known_words(words, model):
    return known(words, model) or known(edits1(word), model) or known_edits2(word, model) or [word]
def correction(word, model):
    return max(known_words([word], model), key=model.get)
def spell_check(text, model):
    return [correction(word, model) for word in words(text)]
# Example usage
text = “Thes is a smple text with an errer.”
WORDS = train(words(open(‘sozlar.xlsx’).read()))
corrected_text = spell_check(text, WORDS)
print(corrected_text)
```

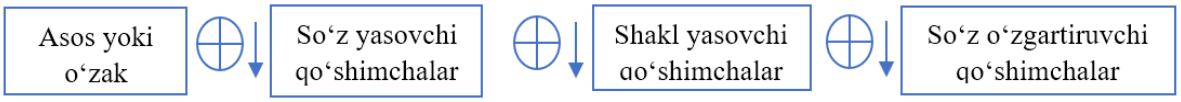
In this example, we will use the ‘spellchecker’ Python package. It finds a list of closed existing words and misspelled words for use in texts and suggests correct options for them. As a result, text filled with correct words is returned. That is, the word base is entered with the file named ‘sozlar.xlsx’. This file contains a list of words from typical texts. Words and their alternatives are selected into the dictionary named ‘WORDS’. The ‘correction’ function finds the best alternative for the entered word. The ‘spell_check’ function checks each word in the text and returns a list of correct spellings.



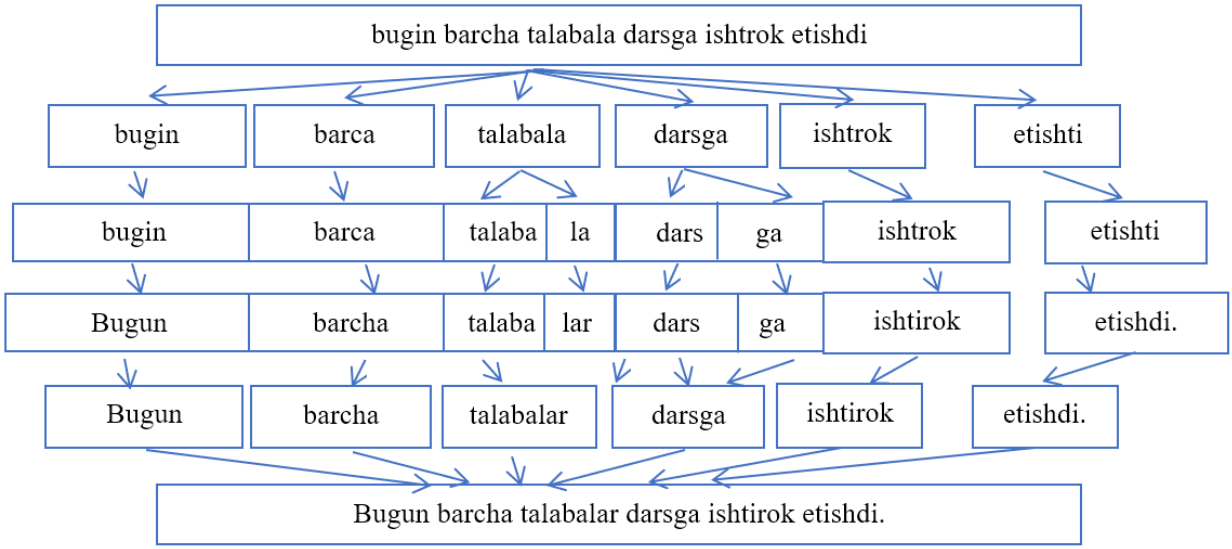
2. UZMORF spell editing process window.

The process of automatic grammatical correction is carried out according to the following algorithms:

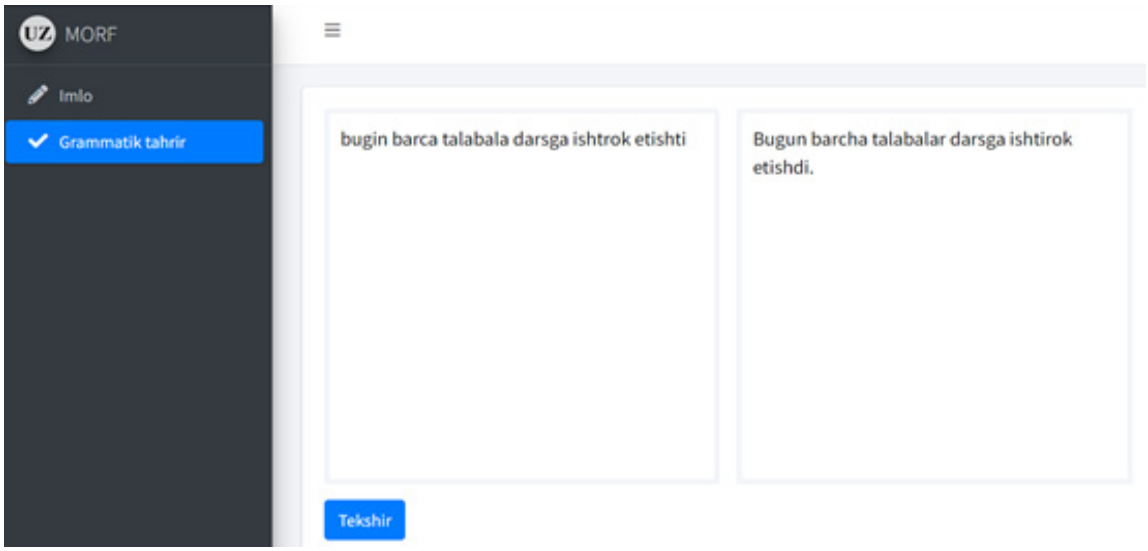
• the entered text is divided into words and suffixes are extracted from the word. Suffixes are attached to the root of the word in a certain sequence, that is:



• each word is checked for availability in the list of word stems available in the database. If it exists, this word is left as it is, otherwise, the grammatical error corresponding to each word is checked from the list of words, and the correct variants of the words written with grammatical error are determined. modified and edited;



• then the user is highlighted the grammatically incorrect words and the system provides the user with an edited text.



3. Grammatical editing process window of UZMORF program

Automatic editing and analysis is one of the main directions of computer linguistics, which arose in connection with the development of the computer text editor (Microsoft word). In the process of automatic editing and analysis, it is assumed that the spelling mistakes of the text entered into the computer will be automatically corrected and the user will be prompted about the error. In the process of automatic editing of the text, in order to make it perfect and high-quality systematic, it is planned to create the possibility of syntactic, semantic and paradigmatic editing of the texts included in it from the orthographic, stylistic and grammatical side.

References

1. Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / — М.: МИЭМ, 2011
2. Nazirova E.Sh., Abidova S.B., Xolmuminov Y.X., va boshq. (2025). Mathematical Model and Algorithm for Grammatical Editing of Given Words in Uzbek Language. Mathematics for Sustainable Industry – Proceedings of ISMI 2024, pp. 175–182 https://www.scopus.com/record/display.uri?eid=2-s2.0-105007504801&origin=recordpage
3. Эшкараева Н.Г, Боймуродов Ф.Ф Разработка и реализация алгоритма стеммативации и генерации словоформ для глаголов узбекского языка. http://qarshidu.uz/source/JURNAL/Aniq_fanlar_5-son_2022.pdf
4. Боймуродов.Ф., Algorithm development and implementation stemmatization and generation of word forms of uzbek language for automatic processing systems texts, https://www.researchgate.net/publication/347995316_ALGORITHM_

DEVELOPMENT_AND_IMPLEMENTATION_STEMMATIZATION_AND_GENERATION_OF_WORD_FORMS_OF_UZBEK_LANGUAGE_FOR_AUTOMATIC_PROCESSING_SYSTEMS_TEXTS

5. Nazirova E.SH, Boymurodov F.F. Tabiiy tilli matnlarni avtomatik qayta ishlash tizimlari tahlili. https://scienceweb.uz/publication/14975
6. Боймуродов Ф.Ф. Стемматизация и генерация словоформ в узбекском языке для систем автоматической обработки текстов. https://scienceweb.uz/publication/14977

IQTISODIYOT&TARAQQIYOT

Ijtimoiy, iqtisodiy, texnologik, ilmiy, ommabop jurnal

Ingliz tili muharriri: Feruz Hakimov

Musahhih: Zokir ALIBEKOV

Sahifalovchi va dizayner: Oloviddin Sobir o‘g‘li

2025. № 6-maxsus son

Mazkur jurnalda maqolalar chop etish uchun quyidagi havolalarga
maqola, reklama, hikoya va boshqa ijodiy materiallar yuborishingiz mumkin.
Materiallar va reklamalar pullik asosda chop etiladi.
El.Pochta: sq143235@gmail.com
Bot: @iqtisodiyot_77
Tel.: 93 718 40 07

Jurnalga istalgan payt quyidagi rekvizitlar orqali obuna bo‘lishingiz mumkin. Obuna bo‘lgach, @iqtisodiyot_77 telegram sahifamizga to‘lov haqidagi ma’lumotni skrinshot yoki foto shaklida jo‘natishingizni so‘raymiz. Shu asosda har oygi jurnal yangi sonini manzilingizga jo‘natamiz.

““Yashil” iqtisodiyot va taraqqiyot” jurnali 03.11.2022-yildan
O‘zbekiston Respublikasi Prezidenti Adminstratsiyasi huzuridagi
Axborot va ommaviy kommunikatsiyalar agentligi tomonidan
№566955 reyestr raqami tartibi bo‘yicha ro‘yxatdan o‘tkazilgan.
Litsenziya raqami: №046523. PNFL: 30407832680027

Manzilimiz: Toshkent shahar, Mirzo Ulug‘bek tumani
Kumushkon ko‘chasi, 26-uy.

Jurnal sayti: https://yashil-iqtisodiyot-taraqqiyot.uz