

УДК 81'322::811.222.8::519.25

ТАҲИЯИ АЛГОРИТМИ ХУЛОСАБАРОРӢ БАРОИ МОДЕЛҲО (АМСИЛАҲО) ОИДИ БА ТАРТИБИ ДУРУСТ ОВАРДАНИ ҶУМЛАИ СОДАИ ТОЧИКӢ ҲАНГОМИ НАВИШТИ НОДУРУСТ

А.А. Қосимов, С.М. Шамсов

Донишгоҳи техникии Тоҷикистон ба номи академик М.С.Осимӣ

Дар мақола масъалаи амсиласозии компютерии қолабҳои маъмули ҷумлаи содаи тоҷикӣ ва таҳияи алгоритми ислоҳ баррасӣ мегардад. Ҳадафи асосӣ таҳия ва татбиқи моделҳои (амсилаҳо) барои намоёндагӣ ва таҳлили сохтори ҷумлаҳои сода мебошад, ки дар онҳо хатоҳои грамматикӣ ва тартиби калимаҳо вучуд дорад. Дар асоси ин моделҳо алгоритми ислоҳ таҳия шуда, ки имкон медиҳад тартиби дурусти вожаҳо дар ҷумлаҳои содаи тоҷикӣ автоматӣ ва дақиқ барқарор намояд. Ин тадқиқот барои рушди барномаҳои коркарди забони табиӣ ва омӯзиши мошинӣ дар забони тоҷикӣ аҳамияти калон дорад.

Калидвожаҳо: амсиласозӣ, қолабҳои маъмули ҷумла, ҷумлаи сода, ислоҳи грамматикӣ, алгоритм, забони тоҷикӣ, коркарди забони табиӣ.

РАЗРАБОТКА АЛГОРИТМА СУММАРИЗАЦИИ ДЛЯ МОДЕЛЕЙ, СВЯЗАННЫХ С ПРАВИЛЬНОЙ ПОСЛЕДОВАТЕЛЬНОСТЬЮ СЛОВ В ПРОСТОМ ТАДЖИКСКОМ ПРЕДЛОЖЕНИИ ПРИ НЕПРАВИЛЬНОМ НАПИСАНИИ

А.А. Косимов, С.М. Шамсов

В статье рассматривается проблема компьютерного моделирования распространённых шаблонов простого предложения таджикского языка и разработки алгоритма их исправления. Основная цель - создание и внедрение моделей (амсил), представляющих структуру простых предложений с учётом типичных грамматических ошибок и неправильного порядка слов. На основе этих моделей разработан алгоритм исправления, который позволяет автоматически и точно восстанавливать правильный порядок слов в простых предложениях таджикского языка. Исследование имеет важное значение для развития программ обработки естественного языка и машинного обучения в таджикском языке.

Ключевые слова: моделирование, распространённые шаблоны предложений, простое предложение, грамматическое исправление, алгоритм, таджикский язык, обработка естественного языка.

DEVELOPMENT OF A SUMMATION ALGORITHM FOR MODELS RELATED TO THE CORRECT SEQUENCE OF WORDS IN A SIMPLE TAJIK SENTENCE WITH INCORRECT SPELLING

A.A. Kosimov, S.M. Shamsov

The article discusses the problem of computer modeling of common patterns of a simple sentence in the Tajik language and the development of an algorithm for correcting them. The main goal is to create and implement models (amsil) representing the structure of simple sentences, taking into account typical grammatical errors and incorrect word order. Based on these models, a correction algorithm has been developed that allows you to automatically and accurately restore the correct word order in simple sentences of the Tajik language. The research is important for the development of natural language processing and machine learning programs in the Tajik language.

Keywords: modeling, common sentence patterns, simple sentence, grammatical correction, algorithm, Tajik language, natural language processing.

Забони тоҷикӣ бо сохтор ва қоидаҳои мураккаби грамматикӣ ва синтаксисӣ фарқ мекунад.[5] Дар забоншиносии ҳисобӣ ва омӯзиши мошинӣ таҳияи моделҳои дақиқи ҷумлабандӣ муҳим аст. Аз ҷумла, амсиласозӣ ё моделсозии қолабҳои маъмули ҷумлаҳо имкон медиҳад, ки сохтори умумии ҷумлаҳо дар барномаи коркарди матн ба таври муассир фаҳмида ва хатоҳои ислоҳ шаванд. [6]

Дар мақола равишҳо ва алгоритмҳои амсиласозӣ ва ислоҳи ҷумлаҳои содаи маъмули тоҷикӣ пешниҳод мегарданд. Амсила (модел) як навъи шакли намунавии грамматикӣ ё синтаксисӣ мебошад, ки сохтори маъмули ҷумлаҳо ифода мекунад. Дар забони тоҷикӣ, ҷумлаҳои сода асосан дорои аъзоҳои пайрав – мубтдо, хабар, муайянкунанда, пуркунанда ва ҳол мебошанд. [6] Муайян кардани дурустии ҷойгиршавии ин аъзоҳо дар ҷумла барои таҳлили грамматикӣ ва ислоҳи хатоҳои муҳим аст. [7]

Барои моделсозии қолабҳои маъмулии ҷумлаҳо ду равиши асосӣ истифода мешаванд:

Қоидабунёд - истифодаи қоидаҳои грамматикӣ ва синтаксисӣ барои сохтани қолабҳо;

Омӯзиши мошинӣ - истифодаи моделҳои омӯзишӣ барои эҷод ва тақмил додани қолабҳо.

Ҳамчунин равиши омехта, ки ҳар ду равишро муттаҳид мекунад, натиҷаҳои беҳтар медиҳад.

Таҳияи алгоритми ислоҳ:

Алгоритми пешниҳодшуда чунин қадамҳоро дар бар мегирад:

- Қабули ҷумлаи вурудӣ (матни нодуруст);
- Таҳлили морфологӣ ва синтаксисӣ барои муайян кардани аъзоҳо;
- Муайян кардани хатоҳои тартиби вожаҳо ва грамматикӣ;
- Барқарор кардани тартиби дурусти вожаҳо дар асоси амсилаҳо;
- Санҷиши натиҷа ва пешниҳоди матни ислоҳшуда.

Барномаи таҳияшуда дар муҳити Python бо истифодаи китобхонаҳои коркарди забони табиӣ (масалан, spaCy, NLTK, Tkinter) сохта шудааст. Моделҳои қоидабунёд бо алгоритмҳои омӯзиши мошинӣ яққо истифода мешаванд, ки дақиқии ислоҳро баланд мебардоранд.

Барнома дорои менюҳои зерин мебошад, ки ҳар яке вазифаҳои мушаххасро барои коркарди матн ва идоракунии луғат таъмин мекунад:

Таҳлили матн: Имконияти таҳлили грамматикӣ ва синтаксисӣ, муайян кардани хатоҳо ва пешниҳоди вариантҳои ислоҳ.

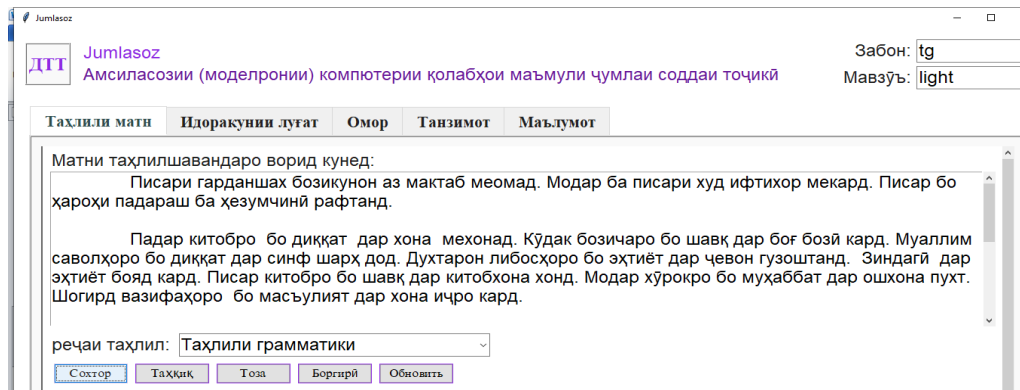
Идоракунии луғат: Мудирияти луғатҳои грамматикӣ бо ҷудокунӣ ба категорияҳои мубтадо, муайянкунанда, пуркунанда, ҳол ва ҳабар.

Омор: Ҳисоботи фаъолияти истифодабаранда бо шикасти вақт (рӯзона, ҳафтаина, моҳона ва солана).

Танзимот: Мураттабсозии параметрҳои барнома барои беҳтар кардани равандҳои таҳлил.

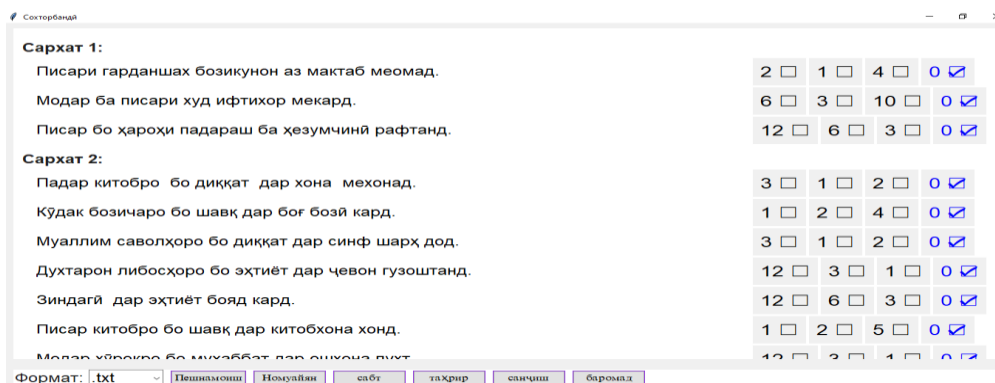
Маълумот: Маълумоти техникӣ ва иттилоот барои истифодабаранда.

Барои таҳлили ҷумлаҳои содаи тоҷикӣ, ки ҳангоми навишти матн, баъзан вақт ҷумларо бо тартиби нодуруст менависанд. Ба мо лозим меояд, ки ҷумларо бо тартиби дуруст орем.



Расми 1 – Ворид кардани матн

Дар ҳолати зер кардани тугмаи “Соҳтор” чунин равзана пайдо мешавад, ки аз чунин қисматҳо ибрат аст, ки дар расми 2. дида мешавад:



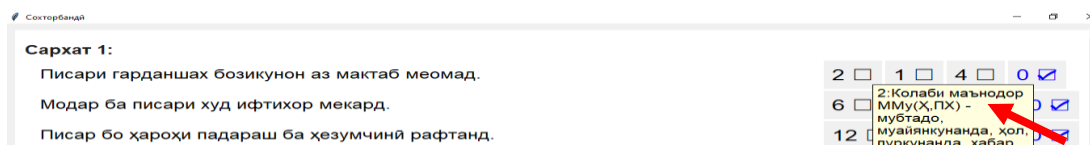
Расми 2 – Сохтани тугмаи “Соҳтор”

Дар равзана дида мешавад матни додашударо ба сархатҳо ҷудо мекунад. Ҳар як сархат дорои миқдори ҷумлаҳо мебошад ва дар бари рости ҳар як ҷумла аз 1 то 3 қолаби наздик меояд. Қолаби 0 барои ба тарзи муқарарӣ (шакли пешина) овардани ҷумла мебошад. Яъне:



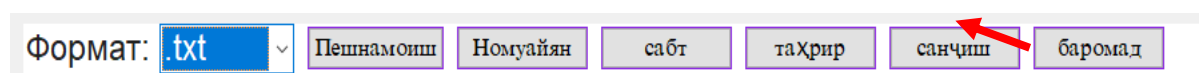
Расми 3 – Қолабҳои наздик ба ҷумла

Ҳангоми интихоби варианти додашуда нишонаи мушакро ба болояш оаварда интихоб кунем, онгоҳ вариант бо намуни қолаб пайдо мегардад.



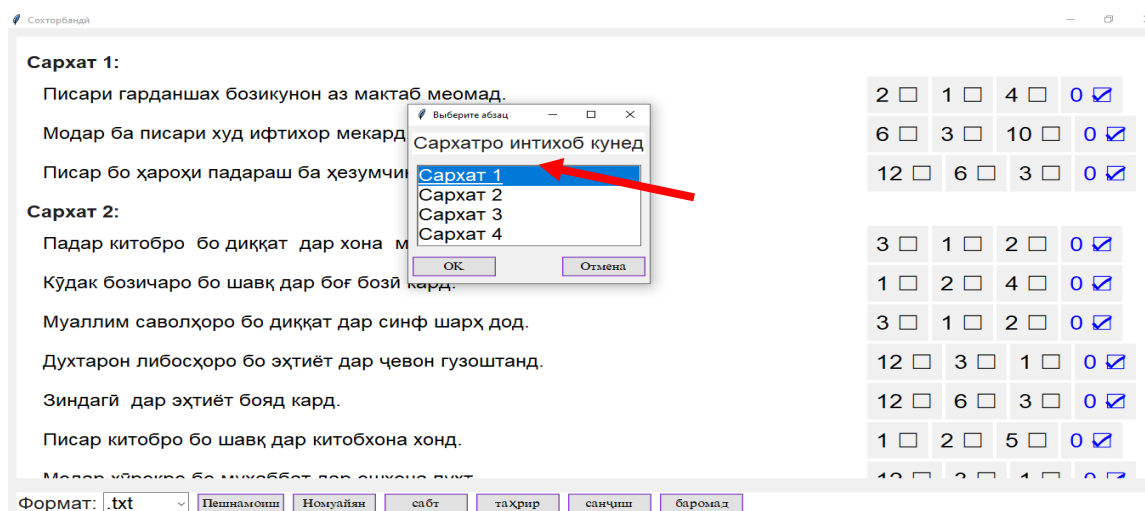
Расми 4 – Интихоби вариант ва намоиши қолаб

Дар вақти интихоби вариантҳо дар поёни равшана 6- то тугма, ки ҳар як тугма вазифаи худро вобаста ба ба тартибдориҳои ҷумла мебошад. Дар вақти интихоби вариантҳо барои ба тартиби дуруст овардани ҷумла дар охир тугмаи “Санҷиш”- ро пахш мекунем.



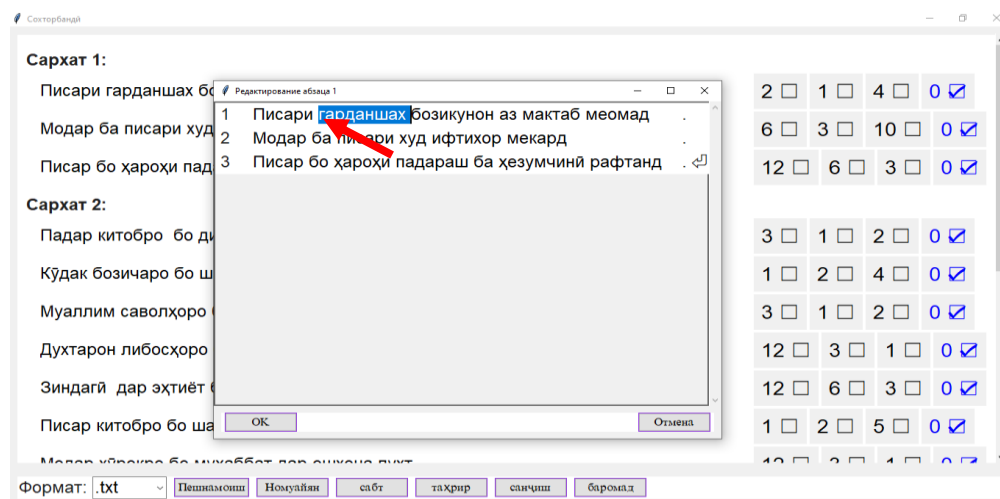
Расми 5 – Тугмаҳо барои иҷрои вазифаҳои махсус оиди ҷумлаҳои интихобишуда

Баъди пахши тугмаи “Санҷиш” ҷумлаҳои интихобкардашуда бо тартиб муайян меояд. Дар ҳолати ба тартиб овардани ҷумлаҳо, агар баъзе ҷумларо каме таҳрир кардан хоҳем барои дурустар шудани маънои ҷумла аз тугмаи “Таҳрир” истифода мекунем. Дар зери тугмаи “Таҳрир” равшана барои таҳрири ҷумлаҳо мебошад. Ҳангоми пахши ин тугма аввал равшана сархатҳоро интихоб мекунанд.



Расми 6 –Интихоби вариант

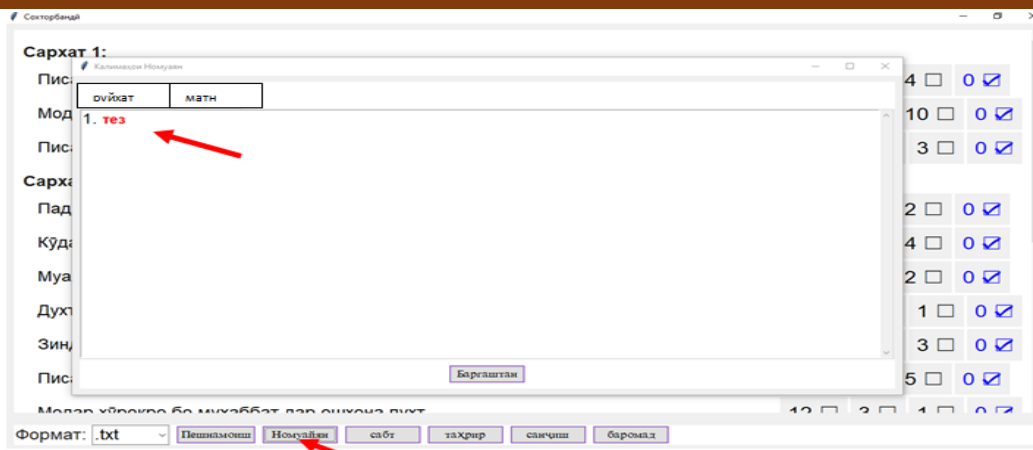
Ҳангоми интихоби “Сархат”-и дилхоҳ разана барои таҳрири ҷумлаҳо алоҳида барои ҳар як сархат пайдо мегардад. Дар ин равшана метавонем калимаҳое, ки мазмунӣ ҷумларо дуруст мекунанд, ворид намоем.



Расми 7 – Майдон барои таҳрири матн

Вақте ки ҳамин корҳоро анҷом додем, баъдан агар дар матн калимаи нав бошад, барнома бо ранги сурх ишорат мекунанд. Ин маънои онро дорад, ки ҳамин калима дар луғати (базаи калима) вучуд надорад. Барои муайян кардани ҳамин гуна калимаҳо дар зери тугмаи “Номуайян” ҷойгир шудааст. Ин равшана асосан ба ду қисм ҷудо мешавад:

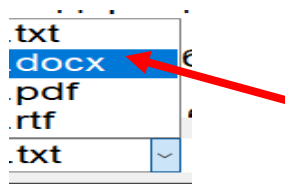
- Руйхат;
- Матн;



Расми 8 – Калимаҳои номуайян

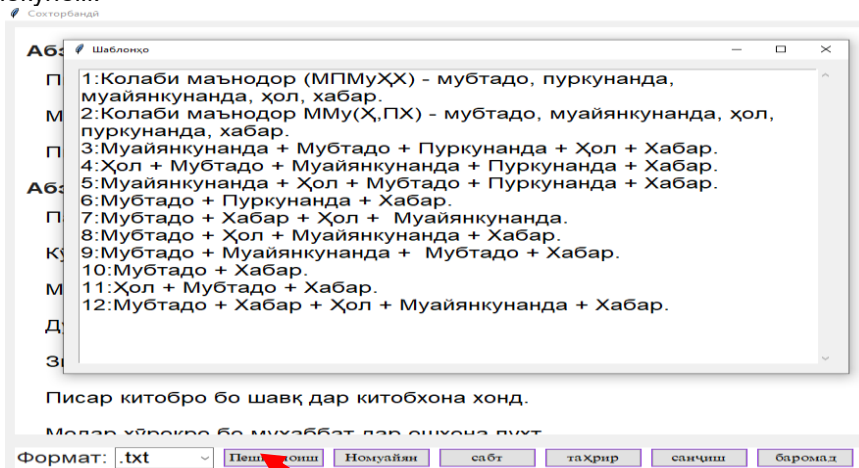
Ин равзана асосан барои муайян кардани калимаҳои, ки дар луғат (манбаи маълумот) нестанд, дар банди рӯйхат нишон дода мешавад. Барои ба майдони асосӣ омадан тугмаи “баргашт”-ро зер мекунем.

Қисмати якумаш “Формат” - аз интихоби намудҳои сабти файл бо форматҳои гуногун аз қабili docx, pdf, rtf, txt:



Расми 9 – Формати файл

Агар ҳамаи ҷумлаҳо робаста аз қолабҳо ба тартиб дарорем, онҳо ҳатман боз ба банди “Қолабҳо” назар кунем. Дар ин равзанаи “Қолабҳо” рӯйхати қолабҳои маънодор мавҷуд аст. Қолаб чист? Яъне ҷумла аз аъзоҳои пайрав иборат аст. Робаста ба ин инҳо барои дурусти кардани ҷумлаҳо бо ин қолаб истифода мекунем.



Расми 10 – Қолабҳои маънодори истифодашаванда

Фақат дар ин равзана қолабҳои маънодор омадааст. Яъне 12-то мебошад. Барои чӣ байни якдигар аломати “+” дорад, барои он ки аъзои пайрав пай дар пай меояд. Инчунин дар равазани сохторбандӣ банди функцияе тартиб дода шудааст, ки ба воситаи зер кардани тугмаи ростӣ мушак пайдо мегардад. Мақсади асосии тартиб додани ин равзана, ки як қисмати ин барнома мансуб мебошад, функцияҳои муҳимро ба анҷом мерасонад. Яъне барои ба таври автоматӣ равона кардани маълумот ба манбаи калимаҳо мебошад. Агар дурустар гуфта гузарем, автоматикунони ҷобаҷо гузори калимаҳо дар манбаи маълумот мебошад. Дар расми 11 намуди равзана оварда шудааст.

Нусхабардорӣ
Ворид кардан
Буридан
Бекор кардан
Такрор кардан
Ҳамаи текстро интихоб кунед
Таъин категория

Расми 11 – Қолабҳои маънодори истифодашаванда

Ҳоло ба таври муфасал оиди ҳар як банди ин равзана пурра дида мебароем:

- Нусхабардорӣ (копировать), яъне барои дилхоҳ қисмати фрагментро интихоб карда, тугмаи чапи мушро зер мекунем, онгоҳ рӯйхати “Нусхабардорӣ” кушода мешавад, ҳангоми интихоби ин банд, аллакай ин қисматро дар хотирааш мегирад ва дар ҷои лозимӣ бурда мегузорад.

Муқарриз: Солиев О.Қ. – д.и.т., дотсент, мудири қабедраи барномасозӣ ва низомҳои зехнии Донишгоҳи байналмилалӣ сайёҳӣ ва соҳибкорӣи Тоҷикистон.

Адабиёт

1. Қосимов А.А. Шамсов С.М. Таҳлили синтаксисии ҷумлаи соддаи паҳншудаи забони тоҷикӣ дар асоси пурқунанда // Конференсияи Ҷумҳуриявии илмӣ-амалии донишҷӯён, магистрантҳо ва аспирантону унвонҷӯён таҳти унвони “Илм – асоси рушди инноватсионӣ”, Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ, шаҳри Душанбе, 2023, с. 105-107.

2. С. Арзуманов, О.Джалолов. Забони тоҷикӣ. - Душанбе “Ирфон”, 1969. -384с.

3. М.А.Исмаилов. Основы автоматизированного морфологического анализа слов таджикского языка -Душанбе,1994. – 150с.

4. АсозодаХ., Кабиров Ш., Анварӣ С. Забон ва адабиёти тоҷик (дастури таълими). - Душанбе, 2005.

5. Грамматикаи забони адабии ҳозираи тоҷик. - Душанбе: Ирфон, 1987.

6. Ғаффоров Р., Хошимов С., Камолиддинов Б. Услубиношӣ. - Душанбе: Маориф, 1975.

7. Каримов У., Абдулазизова С., Забони тоҷикӣ. - Қ. 1-2 - Душанбе: 2003-2004.

8. Арзуманов С. Забони тоҷикӣ. - Душанбе: Маориф, 1986.

9. Аллен Б. Дауни - Think DSP. Цифровая обработка сигналов на Python - Издательство "ДМК Пресс" - 2017 - 160с.

10. Арсак Ж. Программирование игр и головоломок: Пер. 6 франц.- М.: Наука. Гл. ред. физ.-мат. лит., 1990.— 224 С.

11. Бизли Д. Python. Подробный справочник. - Пер. с англ. - СПб.: Символ-Плюс, 2010. - 864 с, ил.

12. Брукс Ф., Мифический человек-месяц, или Как создаются программные системы, М.: Символ-Плюс, 2010.

13. Буйначев, С. К. Основы программирования на языке Python [Электронный ресурс]: учебное пособие / С. К. Буйначев, Боклаг Н. Ю. –2014г.

СВЕДЕНИЯ ОБ АВТОРАХ – МАЪЛУМОТ ДАР БОРАИ МУАЛЛИФОН – INFORMATION ABOUT THE AUTHORS

RU	TJ	EN
Қосимов Абдунаби Абдурауфович	Қосимов Абдунаби Абдурауфович	Kosimov Abdunabi Abduraufovich
Доктор технических наук, доцент	Доктори илмҳои техники, дотсент	Doctor of Technical Sciences, associate professor
Таджикский технический университет имени академика М.С. Осими	Донишгоҳи техникии Тоҷикистон ба номи академик М.С. Осимӣ	Tajik technical university named after academician M.S. Osimi
E-mail: abdunabi_kbtut@mail.ru		
RU	TJ	EN
Шамсов Соҳибҷон Мирзомудинович	Шамсов Соҳибҷон Мирзомудинович	Shamsov Sohijon Mirsomudinovich
докторант PhD	докторант PhD	PhD student
Международный университет туризма и предпринимательства Таджикистана	Донишгоҳи байналмилалӣ сайёҳӣ ва соҳибкорӣи Тоҷикистон	International University of Tourism and Business of Tajikistan
E-mail: sohibjon.sh.93@mail.ru		