

ДОНИШКАДАИ ПОЛИТЕХНИКИИ  
ДОНИШГОҶИ ТЕХНИКИИ ТОҶИКИСТОН  
БА НОМИ АКАДЕМИК М.С. ОСИМӢ ДАР ШАҲРИ ХУҶАНД

КУД 519.25+81'322+811.222.8

  
*Ба ҳуқуқи дастнавис*

**Довудов Гулшан Мирбахоевич**

**ТАҲЛИЛИ МОРФОЛОГИИ КОМПЮТЕРИИ  
КАЛИМАҶОИ ЗАБОНИ ТОҶИКӢ**

**А В Т О Р Е Ф Е Р А Т И**

диссертатсия барои дарёфти дараҷаи илмии номзади илмҳои техникӣ  
аз рӯйи ихтисоси 05.13.11 - Таъминоти математикӣ ва барномавии мошинҳои  
ҳисоббарор, мучтамаъҳо ва шабакаҳои компютерӣ

Душанбе 2018

Кори илмӣ дар Донишкадаи политехникии Донишгоҳи техникии Тоҷикистон ба номи академик М. Осимӣ дар шаҳри Хучанд иҷро карда шудааст.

Роҳбари илмӣ:

Усманов Зафар Чураевич,  
доктори илмҳои физика ва математика,  
академики АИ ҚТ, профессор, мудири  
шӯъбаи моделиронии математикӣ,  
Институти математикаи АИ ҚТ

Муқарризони расмӣ:

Мирзоаҳмедов Фаҳриддин,  
доктори илмҳои техникӣ, профессор, мудири  
шӯъбаи рушди инноватсионии илм,  
Маркази инноватсионии рушди илм ва  
технологияҳои нави АИ ҚТ

Зарипов Саидахмад Асрорович,  
номзади илмҳои физика ва математика,  
и.в. дотсенти кафедраи барномасозии ва  
муҳандисии компютерӣ,  
Донишгоҳи технологии Тоҷикистон

Ташкилоти тақриздиханда:

Муассисаи таълимии таҳсилоти  
касбии олии байнидавлатии  
Донишгоҳи славянии Русияву Тоҷикистон

Ҳимояи диссертатсия санаи 06 июли соли 2018, соати 14:00 дар ҷаласаи шӯрои диссертатсионии 6D.КOA-032 дар назди Донишгоҳи техникии Тоҷикистон ба номи академик М.С.Осимӣ, дар суроғайи шаҳри Душанбе, хиёбони ак. Раҷабовҳо 10 баргузор мегардад.

Бо диссертатсия ва автореферати он дар китобхона ва сомонаи расмии Донишгоҳи техникии Тоҷикистон ([http://ttu.tj/ru/2018/04/06/dovudov\\_g\\_m/](http://ttu.tj/ru/2018/04/06/dovudov_g_m/)) шинос шудан мумкин аст.

Автореферат “ \_\_\_\_\_ ” соли 2018 тавзеъ шудааст.

Хоҳишмандем тақризхоро нисбати автореферат дар ду нусха бо муҳри муассиса ба суроғайи зерин ирсол намоед: 734042, ш. Душанбе, хиёбони академикҳо Раҷабовҳо, 10 тел: (+992 37) 227-37-81, e-mail: [saidaliev.ss@mail.ru](mailto:saidaliev.ss@mail.ru)

**Котиби илмии  
шӯрои диссертатсионии 6D.КOA-032,  
номзади илмҳои техникӣ,**



**Саидалиев Ш.С.**

## ТАВСИФИ УМУМИИ КОР

**Аҳамиятнокии мавзӯ.** Автоматикунони коркарди маълумоти матнӣ яке аз масъалаҳои асосии ҷаҳони муосир ба ҳисоб меравад. Он дар асоси худ ба таҳлили морфологии калимаҳо, ки дар тарҷумаи компютерӣ, тафтиши имлоӣ, таҳлил ва синтези нутқ, муколама бо компютер, гурӯҳбандӣ, фишурданамоӣ, тафсил, чудо кардани калимаҳои калидӣ ва як қатор соҳаҳои дигар истифода мешавад, таъя мекунад.

Аҳамиятнокии амаликунии таҳқиқот оид ба автоматикунони таҳлили морфологии (*ТМ=таҳлили морфологӣ*) калимаҳои забони тоҷикӣ дар Қарори Ҳукумати Ҷумҳурии Тоҷикистон “Дар бораи тасдиқи барномаи истифода ва рушди технологияи иттилоотӣ дар забони тоҷикӣ” аз 06.06.2005 №188 таъкид карда мешавад. Кори илмӣ мазкур айнан барои ҳалли ин масъала равона карда шудааст. Дар он натиҷаи корҳои Ниёзмухаммадов Б., Рустамов Ш., Ниёзӣ Ш., Розенфелд А.З., Расторгуева В.С., Камолиддинов Б., Амонова Ф.Р., Ализода С., Мирзоев Г., Таджиев Д.Т. ва дигарон дар соҳаи морфологияи забони тоҷикӣ истифода шудаанд. Дар он инчунин таҳлилҳои пешоҳангонаи Усманов З.Д., Исмоилов М.А., Исмоилова Р.М., Абдуллаев Ф.А., Назаров Р.С. ва Грашенко Л.А. оид ба омӯзиши ҷанбаҳои алоҳидаи автоматикунони таҳлили морфологии калимаҳои забони тоҷикӣ ба инобат гирифта шудаанд. Яке аз сабабҳои асосие, ки вобаста ба он пешсафони ин ҷода ба ҳалли ниҳоеи масъалаи мазкур расида натавонистанд, ин мавҷуд набудани базаи мукаммали морфҳо мебошад.

Барои ҳалли масъалаи мазкур аз таҷрибаи ғании намояндагони лингвистикаи компютери эронӣ низ ба пуррагӣ истифода бурдан имконнопазир аст. Новобаста аз он ки забони тоҷикӣ як ҷузъи забони форсӣ аст, монетаи асосӣ ин тафовут дар системаи графикӣ мебошад. Дар Эрон алифбои арабиасос ва дар Тоҷикистон алифбои кириллӣ истифода мешавад.

Грамматикаи забони тоҷикӣ дар асоси грамматикаи забони русӣ тартиб дода шудааст. Бинобар ин дар диссертатсия таҷрибаҳои Белоногов Г.Г., Малковский М.Г., Андреев А.М., Демянков В.З., Шуклин Д.Е. ва Ножова И.М. зимни баррасии масъала барои забони русӣ истифода шудаанд.

**Алоқаи кор бо лоиҳаҳои илмӣ.** Кори илмӣ дар доираи барномаи давлатии «Коркард ва таҳқиқоти моделҳои математикӣ барои ҳалли масъалаҳои амалӣ» ГР0116ТJ00533 ҳамчун ҷузъи қисмати «Тартибдиҳии моделҳои математикӣ, алгоритмҳо ва барномаҳо барои коркарди маълумоти матнӣ» иҷро шудааст.

**Мақсад ва муҳтавои таҳқиқот.** Алгоритмизатсияи раванди таҳлили морфологии калимаҳои тоҷикӣ ва амалӣ намудани он дар намуди комплекси барномавӣ.

Барои ноил шудан ба мақсад чунин масъалаҳо матраҳ шуданд:

1. тартиб додани маҷмӯи матнҳои забони тоҷикӣ;
2. тартиб додани протокураи итеративии нимаавтоматӣ (шиносандаи морф)

барои чудо намудани реша ва аффиксҳои калимаҳои маҷмӯи матнҳо;

3. тартиб додани базаи префиксҳо, решаҳо ва постфиксҳои забони тоҷикӣ;

4. пешниҳод намудани тавсифи системавии категорияҳои шаклсоз ва граммемаҳои ҳиссаҳои нутқи забони тоҷикӣ;

5. коркарди таъминоти алгоритмии таҳлилгари автоматии морфологии калимаҳо ва амалӣ намудани он дар намуди комплекси барномавӣ.

**Объекти таҳқиқот** – қонуниятҳои грамматикӣ забони тоҷикӣ.

**Предмети таҳқиқот** – танзими қонуниятҳои омории ташаккули калимаҳои забони тоҷикӣ бо мақсади автоматизатсияи ТМ.

**Методҳои таҳқиқот.** Барои ҳалли масъалаҳои зикршуда методҳои назарияи маҷмӯёҳо, таҳлили системавӣ, кодиронӣ, моделиронии математикӣ, омори математикӣ ва барномарезии баобъектнигаронидашуда барои коркарди воситаҳои барномавӣ истифода шудаанд.

**Навгониҳои илмӣ.** Натиҷаҳои асосии диссертатсия нав буда, чунин ифода мегарданд:

- пас аз коркарди маҷмӯи матнҳо дар ҳаҷми 59 344 883 калима, базаи бузурги морфҳои забони тоҷикӣ ташкил карда шуд, ки аз 81 префикс, 76 539 реша ва 128 760 постфикс иборат аст;

- таснифи намуди аффиксҳо (шаклсоз, калимасоз ва иборасоз) ва калимаҳо пешниҳод шуд;

- кодиронии мавқеии калимаҳои забони тоҷикӣ коркард шуд;

- алгоритми табдили калимаҳои иборасоз ба ибораҳо пешниҳод шуд;

- таъминоти алгоритмии ТМ калимаҳои забони тоҷикӣ коркард шуд.

**Аҳамияти назариявӣ** диссертатсия дар он зуҳур меёбад, ки методҳои барои он коркардшудаи таҳлили морфологӣ дар забонҳои дигари гурӯҳи эронӣ истифода шуда метавонанд.

**Аҳамияти амалии кор.** Таҳлилгари морфологии дар доираи диссертатсия коркардшуда дар Маркази миллии патенту иттилоотии Вазорати рушди иқтисод ва савдои Ҷумҳурии Тоҷикистон (ВРИваС ҚТ) ба намуди захираи иттилотӣ таҳти рақами ЗИ-03.2.220ТJ аз 20.12.2011 ба қайд гирифта шудааст. Он имкониятҳои васеъро барои коркарди автоматии матнҳои забони тоҷикӣ фароҳам меоварад.

Дар асоси ТМ пакетҳои забонӣ барои тафтиши имлоии калимаҳои забони тоҷикӣ дар OpenOfficeOrg ва Microsoft Office тартиб дода шудааст, ки ҳамчун захираҳои иттилоотии ЗИ-03.2.222ТJ аз 11.01.2012 ва № 4201200235 аз 04.10.2012 ба қайд гирифта шудаанд. Пакетҳои мазкур дар аксари муассисаҳо ва корхонаҳои Ҷумҳурии Тоҷикистон истифода мешаванд.

**Нуктаҳои, ки барои химоя бароварда мешаванд:**

- коркарди шиносандаи морф – протседураи нимаавтоматии итератсионӣ барои муайян намудани решаҳо ва аффиксҳои калимаҳои забони тоҷикӣ;

- тартиб додани базаи морфҳои забони тоҷикӣ;

- таснифи намуди аффиксҳо ва калимаҳои забони тоҷикӣ;
- коркарди кодиронии мавқеии калимаҳои забони тоҷикӣ;
- табдили калимаҳои иборасоз ба порчаи ҷумла;
- коркарди таҳлилгари автоматии морфологии калимаҳои забони тоҷикӣ.

**Саҳеҳият ва асоснокӣ** натиҷаи асосии диссертация - коркарди *таҳлилгари автоматии морфологии* калимаҳои забони тоҷикӣ, бо татбиқ ва ташаккули методҳои самараноки эътирофшудаи гибридии барои иҷрои таҳқиқоти лингвистикӣ ва имконияти дар сатҳи 92,2% дуруст ҷудо намудани калимаҳои забони тоҷикӣ ба морфҳо, ки бо роҳи таҷрибаи ҳисобӣ аниқ карда шудааст, кафолат дода мешавад.

**Саҳми муаллиф.** Гузориши масъала бо ҳамчоягии роҳбари илмӣ амалӣ шудааст. Натиҷаҳои асосии кори диссертационӣ аз тарафи муаллиф мустақилона ба даст оварда шудааст.

**Татбиқи кор.** Натиҷаҳои ҷудогонаи кор дар конференсияҳо ва семинарҳои зерин пешниҳод ва муҳокима гардидаанд:

- “Recent Advances in Slavonic Natural Language Processing”. Донишгоҳи Масарик, шаҳри Брно, Чехия, 2010-2012;

- “Системаҳои иттилоотии амалӣ: масъалаҳои моделиронӣ, коркард ва истифодаи он дар мамлакатҳои тараққикардаи тоҷа”. Хучанд, 29-30 июни 2012;

- “Language Technology for Normalisation of Less-Resourced Languages SALT MIL 8 - AFLAT 2012”. Истамбул, Туркия 22.05.2012;

- семинарҳои илмӣ-таҳқиқотӣ оид ба лингвистикаи компютерӣ дар Институти математикаи АИ ҶТ, Донишгоҳи технологияи Тоҷикистон ва Донишгоҳи славянии Тоҷикистону Руссия дар солҳои 2013-2017;

- конференсияи илмӣ-амалии байналмилалӣ “Нақши технологияҳои иттилоотию коммуникатсионӣ дар рушди инноватсионии Ҷумҳурии Тоҷикистон”. Душанбе, 17-18 ноябри 2017.

**Интишороти асосӣ.** Аз рӯи мавзӯи диссертационӣ 19 кор: 2 монография, [138, 139], 13 мақола, [140-152] ва 4 шаҳодатнома дар бораи бақайдгирии давлатии захираи иттилоотӣ нашр шудаанд. Аз онҳо 10 мақолаи илмӣ, дар нашрияҳои аз тарафи КОА –и назди Президенти Ҷумҳурии Тоҷикистон тавсияшуда, нашр гардидаанд.

**Соҳтор ва ҳаҷми кор.** Рисолаи илмӣ аз мундариҷа, панҷ боб, хулоса ва рӯйхати адабиёти истифодашуда иборат аст. Муҳтавои диссертатсия дар 120 саҳифа дарҷ гардида, 38 ҷадвал ва 41 расмро дар бар мегирад.

## МУНДАРИЦАИ ДИССЕРТАТСИЯ<sup>1</sup>

Дар муқаддима элементҳои сохтори асосии матни диссертатсия дар мувофиқат бо ГОСТ Р 7.0.11-2011 баён гардидааст.

**Боби 1** ба ташкили базаи нисбатан пурраи морфҳо дар асоси коркарди файлҳои матнӣ бо истифодаи барномаи шиносандаи морф (*MR*) бахшида шудааст. *MR*, ки бо  $MR(n_1, n_2, n_3)$  ифода мегардад, аз як тараф, протседураи нимаавтоматист, ки дар асоси коркарди матн базаи морфҳоро ташкил медиҳад ва аз тарафи дигар, “ташхиси дастӣ”-ест, ки калимаро ба морфҳо тақсим мекунад. Рақамҳои  $n_1$ ,  $n_2$  ва  $n_3$  мувофиқан ифодакунандаи миқдори префиксҳо, решаҳо ва постфиксҳои дар базаи *MR* буда, пеш аз иҷрои иттератсияи навбатӣ мебошанд.

Тақсими автоматии калима ба морфҳо аз тарафи компютер иҷро мешавад. Барномаи компютерӣ дар асоси алгоритми додасуда калимаро ба морфҳо тақсим мекунад. Вобаста ба ҳаҷми базаи морфҳои дар барномаи *MR* истифодашаванда, калимаи дилхоҳ метавонад, ё ба се қисм тақсим карда шавад ва ё чунин тақсимот имконнопазир бошад.

Чудокунии морфҳо мумкин аст, натиҷа надихад. Азбаски калимаи таҳлилшаванда аз ҷиҳати имлоӣ дуруст аст, сабаби ба морфҳо тақсим карда натавонистани калима дар база мавҷуд набудани реша ва ё аффикси мувофиқ мебошад. Дар ин ҳолат калима барои таҳлили морфологӣ ба эксперт раван карда мешавад.

Қисмати муайяни кор – таҳлили калимаҳои ба таври автоматӣ ба морфҳо тақсимнашуда ба дӯши эксперт воғузошта мешавад. Эксперти забоншинос, дар сурати муайян намудани мансубияти калима ба забони тоҷикӣ, онро ба морфҳо тақсим мекунад. Морфҳои муайяншуда барои тақмили базаи морфҳо истифода мешаванд. Агар калима ба забони тоҷикӣ мансуб набошад, он гоҳ он сарфи назар карда мешавад (дар оянда таҳлил намешавад).

Барои амалӣ намудани таҳқиқоти оморӣ дар § 1.2 маҷмӯи матнҳои забони тоҷикӣ бо андозаи 59344883 калимаҳо (273734 калимаҳои уникалӣ) мураттаб шудаанд.

Барои ташкили базаи префиксҳо дар §§ 1.3 ва 1.4 –и диссертатсия методи комбинаторӣ-оморӣ (К-О) пешниҳод мешавад, ки моҳиятан чунин аст. Маҷмӯи префиксҳои  $\{PR\}$  забони тоҷикӣ дар чунин намуд ифода мешаванд:

$$\{PR\} = \bigcup_{k=1}^{k=3} \{PR(k)\}, \quad (1.2)$$

яъне, ҳамчун ҳамчояшавии назариявӣ-маҷмӯии се зермаҷмӯи  $\{PR(k)\}$ ,  $k = 1, 2, 3$ . Элементҳои зермаҷмӯи  $\{PR(1)\}$  префиксҳои яктаркиба(сода) буда, префиксҳои зермаҷмӯъҳои  $\{PR(2)\}$  ва  $\{PR(3)\}$  префиксҳои дутаркиба ва сетаркибаанд, ки онҳоро шартан префиксҳои мураккаб меномем.

---

<sup>1</sup> Рақамгузори бобҳо, зербобҳо, формулаҳо, ҷадвалҳо, расмҳо ва ғ. дар автореферат мувофиқи рақамгузори онҳо дар диссертатсия истифода мешавад.

Чунин фарзияҳоро қабул мекунем:

– *рӯйхати префиксҳои сода ба, бар, бе, би, бо, боз, бу, во, дар, ма, ме, на, но, то, фар, фур, ҳам, ҳаме, ҳа* мукамал аст, яъне, ғайр аз ин 19 префикс дар забони тоҷикӣ дигар префикси сода мавҷуд нест;

– *префикси мураккаби дилхоҳ аз ҳамҷояшавиши префиксҳои содаи гуногун сохта мешавад*: ду префикс барои префиксҳои дутаркиба ва се префикс барои префиксҳои сетаркиба. Дар таркиби префикси мураккаб префикси содаи такроршаванда мавҷуд нест.

Зермаҷмӯҳои  $\{PR(2)\}$  ва  $\{PR(3)\}$  мувофиқан на зиёда аз 342 ( $= 19 \cdot 18$ ) префикси дутаркиба ва 5814 ( $= 19 \cdot 18 \cdot 17$ ) префиксҳои сетаркибаро дар бар мегиранд. Элементҳои ба таври формалӣ бадастомадаи зермаҷмӯҳои  $\{PR(2)\}$  ва  $\{PR(3)\}$ -ро *префиксҳои виртуалӣ* меномем. Рӯйхати пурраи онҳо барои дар оянда муайян намудан ва тартиб додани рӯйхати префиксҳои *ҳақиқӣ* ва *ғайриҳақиқӣ* истифода мешаванд.

Сипас бо ёрии барномаи компютерӣ ва истифодаи зермаҷмӯи  $\{PR(1)\}$  - префиксҳои сода ва зермаҷмӯҳои  $\{PR(2)\}$  ва  $\{PR(3)\}$  - префиксҳои виртуалӣ ҳар як калима ба мақсади чудо намудани префикси ҳақиқӣ ташхис карда шуд. Ҳангоми ташхис чунин қоида ба инobat гирифта шуд:

**Қоида.** *Агар баъди аз калима чудо намудани префикси виртуалӣ пайдарпайиши боқимондаи символҳо калимаи маънодор бошад, он гоҳ ин префикс - префикси ҳақиқӣ ҳисобида мешавад.*

Раванди шинохти префиксҳо аз тарафи эксперт иҷро карда шуд. Дар натиҷаи коркарди маҷмӯи матнҳо бо ёрии  $MR$ , 53 префикси дутаркиба (барма-, барме-, барна-, ..., ҳамефар-, ҳамефур-) ва 9 префикси сетаркиба (барнаме-, бознаме-, вонаме-, дарнаби-, дарнаме-, намебар-, намедар-, намефар-, намефур-) бо басомади вохӯриашон муайян шуданд. Ҳамин тавр, бо истифодаи методи К-О 81 префикси ҳақиқии забони тоҷикӣ ( $19+53+9$ ) муайян карда шуд.

Базаи префиксҳои ҳосилшуда имконият дод, ки дар § 1.4.1 *тақсимооти префиксҳо аз рӯи дарозӣ, биграммаҳои аввала, ҳарфҳои аввал* муайян карда шавад.

**Базаи постфиксҳо.** Татбиқи методи К-О, ки барои ташкили базаи префиксҳо бомуваффақият истифода шуд, барои ташкили базаи постфиксҳо имконнопазир аст (шумораи постфиксҳои сода – 113, миқдори максималии постфиксҳои сода дар таркиби постфикси мураккаб – 8, барои нигоҳ доштани комбинатсияи постфиксҳои сода ва мураккаб, 2, 3, ...8-таркиба,  $4.16 \cdot 10^8$  Гб хотираи компютер лозим аст). Аз ин сабаб, барои ташкили базаи постфиксҳо *протсекураҳои итератсионӣ* (қадам ба қадам зиёд намудани базаи постфиксҳо) истифода гардид. Дар аввал маҷмӯи матнҳои иборат аз 3800 саҳифа, ки дорои 1 540 019 калимаҳои онд, коркард шуданд (§ 1.5.). Ин амалро  $MR(66, 6000, 270)$ , ки базаи маълумоти он аз 66 префикс (миқдори ниҳоии префиксҳо баъдтар ба даст омад), 6000 реша ва 270 постфикс иборат аст, иҷро намуд. Бар шумори постфиксҳо 113 постфиксҳои сода: **й**,

ад, ам, амон, анд, аст, ат, атон, аш, ашон, ву, ед, ем, етон, и, ро, й, ст, у, ю, яд, ям, ямон, янд, яст, ят, ятон, яш, яшон, а, ак, ан, анда, андар, бон, бор, вй, ванд, вар, ваш, вум, вода, вон, вор, вот, вум, гй, гар, гах, гин, гон, гона, гор, гох, гун, гуна, дон, е, ё, ев, ева, евич, евна, ён, ёнд, зор, ид, ин, ина, истон, иш, када, лох, мй, манд, монанд, навард, но, нок, о, ов, ова, ович, овна, ок, он, она, онд, ор, осо, от, пона, сон, сор, стон, тар, то, тоб, ум, фом, ҳак, ҳо, хон, чй, ча, чак, чон, хот, юм, шан, як, ян, янда ва баъзе постфиксҳои мураккаби қаблӣ дохил шуданд.

Пас аз иҷрои 10 итератсия, ки дар ҳар яки он тақрибан 150000 калима коркард шуд, миқдори постфиксҳо то 2533 ва миқдори решаҳо то 25323 зиёд гардид. Миқдори префиксҳо тағйир наёфт. Версияи нави  $MR(66, 25323, 2533)$  ҳосил гардид.

**Постфиксҳои парадигмаофар (п-о).** Бо мақсади тавсифи миқдори бузурги постфиксҳои забони тоҷикӣ ифодаи постфикси дилхоҳи  $PS$  ба намуди конкатенатсияи

$$PS = PS^1 \oplus PS^2 \quad (1.3)$$

суффикси  $PS^1$  ва бандҳои  $PS^2$ , яъне, ба намуди постфиксҳои п-о ва п-с, мувофиқи мақсад аст. Муайян шуд, ки постфиксҳои п-о  $PS^1$  дар таркиби худ то 5 постфикси сода дошта метавонанд.

**Постфиксҳои парадигмасоз (п-с).** Таҳлилҳо нишон доданд, ки аз 2533 постфиксҳо 127 постфиксро ҷудо кардан мумкин аст, ки дар ташкили парадигмаи постфиксҳои ба маҷмӯи  $PS^1$  тааллуқдошта иштирок меkunанд. Пас аз таҳлил маълум гардид, ки рӯйхат номукамал буда, аз ҳисоби постфиксҳои п-с, ки дар байни 127 постфикс мавҷуд нестанд, онро то 344 адад зиёд кардан мумкин аст.

***MR-и такмилдодашуда.*** Омӯзиши структураи 2533 постфиксҳои версияи охири  $MR$  имконияти муайян намудани 657 элементҳои маҷмӯи  $PS^1$  ва 127 элементҳои маҷмӯи  $PS^2$ -ро (баъд то 344 адад зиёд карда шуд) фароҳам овард. Дар асоси ду маҷмӯъ 46263 постфикс ҳосил карда шуд, ки ҳамаи он аз тарафи эксперт ҳамчун постфикси ҳақиқӣ эътироф гардид.

Базаи постфиксҳои тартибдодашуда барои истифода дар версияи нави  $MR(66, 25323, 46263)$  ноқулай буда, боиси суст гаштани суръати коркарди автоматии маълумот мегардад. Аз ин сабаб, муқаррар гардид, ки дар база 657 элементи маҷмӯи  $PS^1$  нигоҳ дошта шуда, функсияи  $MR$  аз ҳисоби илова кардани модули шинохти автоматии постфиксҳои  $PS^2$  дар асоси маълумот оид ба 344 элементи маҷмӯи  $PS^2$  ва қоидаҳои пайвастшавии постфиксҳои сода васеъ карда шавад.

Версияи такмилдодашудаи  $MR$  ҳамчун  $MR^*(66, 25323, 657)$  ишора шуда,  $MR$ -и такмилдодашуда номида шуд, ки он ҳам ба ташкили коллексияи мукаммали морфҳои забони тоҷикӣ равона гардидааст.

**Қонуниятҳои омории маҷмӯи постфиксҳо.** Қонуниятҳои омории дар § 1.5.1



овардашаванда дар асоси таҳлили маҷмӯи постфиксҳо, ки бо истифодаи  $MR^*(66,25323,657)$  аз маҷмӯи матнҳои андозаашон 55184508 калима ҷудо карда шудаанд, ба даст омадаанд. Коркарди маълумот бо чунин тарз сурат гирифт. Маҷмӯъ ба 10 зермаҷмӯъ, ки ҳар яке дорои 5-5,1 миллион калима мебошад, тақсим гардид. Зермаҷмӯъҳо пай дар пай коркард шуданд. Дар аввал маълумоти матнӣ ба луғати басомадии калимаҳо табдил дода шуд. Сипас ҳар як калима бо ёрии  $MR^*$  ба морфҳо ҷудо гардид. Агар протсекураи ҷудо намудани калима ба морфҳо бомуваффақият анҷом ёбад,  $MR^*$  ба таҳлили калимаи оянда мегузарад. Дар ҳолати бенатича анҷом ёфтани ҷудокунии калима ба морфҳо, калима ишора мешавад. Пас аз анҷоми раванди коркарди зермаҷмӯъ калимаҳои таҳлилнашуда барои таҳлил ба эксперт пешниҳод мешаванд. Эксперт калимаҳоро бо морфҳо ҷудо мекунад. Морфҳои (префиксҳо, решаҳо ва постфиксҳо) дар натиҷаи таҳлили экспертгон ҳосилшуда, дар сурати дар база мавҷуд набудан, барои мукамал намудани базаи морфҳои  $MR^*$  истифода мешаванд.

Дар итератсияи оянда версияи такмилшудаи  $MR^*$  барои таҳлили зермаҷмӯи минбаъда истифода мешавад. Пас аз иҷро намудани даҳ итератсия, миқдори постфиксҳои сода то 125 (аз ҳисоби 12 элементҳои нав **-акак, -ас,-вона, -вора, -гора, -ёна, -ол, -ос, -со, -ур, -ҳакак, -якак**) расид ва  $MR^*(81,65422,2847)$  -и ниҳой ҳосил гардид, ки имконияти шинохти 122 729 постфиксҳои ҳақиқиро дорост. Сипас, баъди коркарди маҷмӯи матнҳои иборат аз 55 184 508 калимаҳо (240 208 калимаҳои уникалӣ) маълум гардид, ки 51 %-и ҳамаи калимаҳо ва 21 % -и калимаҳои уникалӣ дар таркиби худ постфикс надоранд. Илова бар ин, муайян гардид, ки постфиксҳои дар таркибашон зиёда аз 5 постфикси сода дошта дар 100 000 калима на зиёда аз як маротиба мушоҳида мешаванд. Қайд кардан зарур аст, ки ба постфиксҳои зинаи ҳаштуми мураккабӣ постфиксҳои намуди **ишмандтаринҳояшонрову** бо назардошти ҷудо намудани он ба **иш-манд-тарин-хо-яшон-ро-ву** дохил мешаванд.

**Базаи решаҳо.** Бо ёрии  $MR^*$  дар § 1.6 базаи бузурги иборат аз 65 422 решаи калимаҳои забони тоҷикӣ тартиб дода шуд.

**Масъала оид ба мукаммалии базаи морфҳо.** Барои баррасии масъалаи мазкур маҷмӯи матнҳои газетаву маҷаллаҳо, ки аз 877 саҳифа иборат аст (503 725 калимаҳо, 47 224 калимаи уникалӣ), бо ёрии  $MR^*(81,65422,2847)$  коркард шуд.

Муайян гардид, ки 0.356 % аз миқдори умумии калимаҳо ва 1.8 % калимаҳои уникалӣ бо таври автоматӣ ба морфҳо ҷудо карда нашудаанд. Сабаби он мавҷудияти калимаҳои хоричӣ, исмҳо хос ва хатои имлоӣ дар матнҳо аст. Дар ин ҳолат калимаҳои таҳлилнашуда барои таҳлил ба эксперт равона гардиданд. Эксперт хатоҳои мавҷударо ислоҳ намуда, калимаҳоро таҳлил намуд. Морфҳои нави ҳосилшуда ҳиссаи ночизро аз миқдори умумии ҷавобҳои рад ташкил медиҳанд. Ҳамин тавр, намунаи нисбатан мукаммали  $MR^*(81,66108,2862)$  ҳосил гардид. Дар  $MR^*(81,66108,2862)$  постфиксҳои шумораашон то 2862 зиёдгаштаи маҷмӯи  $PS^1$

имконияти шинохти 125 020 постфиксро доранд.

**Тасаввурот оид ба вижагиҳои масъала** бо роҳи коркарди тасодуфии 10 файли матнии андозаашон 305000-496000 калима, ки аз газтаву маҷаллаҳо гирифта шудаанд, ҳосил гардид. Яке аз ин даҳ файл ба таври тасодуфӣ интихоб гардида, бо истифодаи  $MR^*$  (81,66108,2862) коркард шуд. Морфҳои нави ошкоршуда (решаҳо, префиксҳо ва постфиксҳои маҷмӯи  $PS^1$ ) ба базаи  $MR^*$  илова шуданд. Ҳамин тавр, барнома ба версияи нисбатан мукаммал табдил ёфт. Сипас, яке аз нух файлҳои боқимонда ба таври тасодуфӣ интихоб шуда, бо истифодаи  $MR^*$  - и ҳосилшуда коркард гардид. Морфҳои нави ошкоршуда барои тартиб додани версияи нави  $MR^*$  истифода шуданд.

Натиҷаи коркарди ҳамаи 10 файл дар ҷадвали 1.28 нишон дода шудааст. Қайд кардан зарур аст, ки дар сутунҳои 9 ва 12 маълумот танҳо оид ба постфиксҳои маҷмӯи  $PS^1$  оварда шудааст.

Ҷадвали 1.28. - Басомади пайдо шудани морфҳои нав

№ файл	Миқдори калимаҳо	Миқдори калимаҳои уникалӣ	Миқдори ҷавоби рад	Миқдори решаҳои нав	Решаҳои сода	Решаҳои мураккаб	Миқдори префиксҳои нав	Миқдори постфиксҳои нав	% префиксҳо	% решаҳо	% постфиксҳо
1	496724	49637	800	600	150	450	0	18	0	1,208	0,036
2	424712	48762	1016	829	62	767	0	11	0	1,7	0,022
3	310511	12812	716	525	63	462	0	8	0	4,098	0,062
4	342303	11688	545	381	49	332	0	5	0	3,26	0,043
5	343562	12407	531	381	60	321	0	6	0	3,07	0,048
6	307297	13084	488	386	43	343	0	3	0	2,95	0,023
7	328127	13034	315	226	20	206	0	4	0	1,73	0,03
8	320400	13478	205	133	16	117	0	5	0	0,99	0,037
9	305674	12225	253	176	13	163	0	6	0	1,44	0,049
10	477340	37396	910	742	12	730	0	7	0	1,984	0,018
	3656650	224523	5779	4379	488	3891	0	73	0		

Дар сутуни аввали ҷадвал рақами тартибии 10 файли коркардшуда оварда мешавад; дар сутуни дувум шумораи умумии калимаҳои дар файл мавҷуда нишон дода мешавад; дар сутуни сеюм – миқдори калимаҳои уникалӣ дар байни шумораи умумии калимаҳои таҳлилшуда; дар сутуни чорум маълумот оид ба миқдори калимаҳои таҳлилнашуда, яъне, ҳолатҳое, ки барнома калимаро ба морфҳо ҷудо карда натавонист, оварда мешавад. Сабаби ин мавҷудияти морфи нав (префикс, реша ва ё постфикс) дар таркиби калима аст. Айнан чунин калимаҳо ба эксперт-забоншинос барои таҳлил пешниҳод карда мешаванд. Эксперт натиҷаи таҳлиро ба сутунҳои 5-9 дохил мекунад. Дар сутунҳои 10-12 ҳиссаи морфҳои ошкоршуда нисбат ба калимаҳои уникалии файлҳо оварда мешавад. Сутуни 11 оид ба натиҷаҳои ниҳони коркарди маҷмӯи тасодуфии аз 3 656 650 иборатбуда маълумот медиҳад. Пас аз коркарди маълумот ягон префикси нав ошкор нашуд, лекин 4379

решаҳои нав ва 73  $PS^1$  - постфиксҳо муайян шуданд. Дар натиҷа, версияи нави  $MR^*$  (81, 70487, 2935) ҳосил гардид, ки 128 760 постфиксро шинохта метавонад.

Ҳамин тавр, базаи ибтидоӣ аз 81 префикс тағйир наёфт. Вобаста ба ин фарзия пешниҳод мешавад, ки базаи префиксҳо ба таври кофӣ мукамал аст. Бо зиёд шудани ҳаҷми матнҳои коркардшаванда, шумораи  $PS^1$  -постфиксҳои ошкоршаванда тамоюли наздикшавиро инъикос намекунанд (нигаред ба сутунҳои 9 ва 12). Нисбат ба решаҳо низ ҳолат чунин аст: сутунҳои 5 ва 11 имкон намедиханд, ки ақалан баҳои тахминии мукаммалии луғати решаҳо, ки имконияти таҳлили морфологии калимаҳои тоҷикиро медиҳанд, муайян намоем.

Натиҷаҳои дар поён овардашуда, пеш аз ҳама, ба калимаҳои сода тааллуқдоранд, ки ба яке аз чор структураҳои имконпазир ( $R$  - калима аз реша иборат аст,  $PR \oplus R$  - калима аз префикс ва реша иборат аст,  $PR \oplus R \oplus PS$  - калима аз префикс, реша ва постфикс иборат аст,  $R \oplus PS$  - калима аз реша ва постфикс иборат аст) мутааллиқанд.

Ҷадвали 1.31. - Тақсимоти басомади вохӯрии структураҳои калима бо %

Миқдор	$R$	$PR \oplus R$	$PR \oplus R \oplus PS$	$R \oplus PS$	Ҳамагӣ
Калимаҳои уникалӣ	18,87	1,58	8,41	71,14	100
Калимаҳо	49,99	1,07	4,27	44,67	100

Аз ҷадвали мазкур бармеояд, ки миқдори зиёди калимаҳои тоҷикӣ (94.66 % - дар байни калимаҳо ва 90.01% - дар байни калимаҳои уникалӣ) сохтори бидуни префикс доранд.

Дар **боби 2** фаҳмиши лозимӣ оид ба структураи калимаҳои забони тоҷикӣ ва маҷмӯи граммемаҳои калима тартиб дода мешавад. Граммемаҳои калима қиматҳои мушаххаси категорияҳои грамматикӣ калимаро инъикос мекунанд, ки он заминаро барои ташаккули таъминоти алгоритмии таҳлили морфологӣ фароҳам меоварад.

Шиносандаи морфи дар боби 1 тавсифшуда имконияти ҷудо кардани калимаи дилхоҳро ба реша, асос ва аффиксҳо (префиксҳо ва постфиксҳо) дорад. Аммо ин кифоя нест. Барои ноил шудан ба мақсади ниҳоии таҳлили морфологӣ донишмандони ҳиссаи нутқи реша, асос ва маҷмӯи граммемаҳои мувофиқи калима лозим аст.

Талаботи объективии таҳлили морфологӣ моро водор намуд, ки бар ивази даҳ ҳиссаи нутқи маъмулӣ, **нуҳ ҳиссаҳои мустиқили нутқ** – *исм, сифат, шумора, ҷонишин, феъл, масдар, сифати феълӣ, феъли ҳол, зарф* ва **ҳафт ҳиссаҳои номустақили нутқ** – *пешоянд, пасоянд, пайвандак, ҳиссаҷа, нидо, калимаи тақлиди овозӣ, нумеративро* истифода барем. Дар боби 2 тавсифи системавии ҳиссаҳои нутқ бо категорияҳои шаклсозии грамматикӣ онҳо ва рӯйхати имконпазири қиматҳои грамматикӣ ин категорияҳо оварда мешавад.

Ҳиссаҳои мустиқили нутқ системаи мукаммали категорияҳои грамматикӣ доранд. Ҳар як категория бо маҷмӯи граммемаҳои худ, яъне, бо қимати граммема-

ҳои шаклсоз ифода мегардад. Граммемаҳо бо воситаи морфемаҳои шаклсоз, ки ба решаи (асос) калима пайваст шуда, калимаи нав месозанд, ифода мегарданд. Мафҳуми “шаклсоз” ифодакунандаи он аст, ки решаи (асос) калима ва калимаи аз он сохташаванда ба як ҳиссаи нутқ тааллуқ дорад. Бо маънои дигар, ҳиссаи нутқ нисбат ба амали пайваст кардани аффикси шаклсоз ба реша инвариантӣ мебошад.

Ҳафт ҳиссаи нутқи боқимонда (пешоянд, пасоянд, пайвандак, ҳиссача, нидо, калимаи тақлидиовозӣ, нумератив) ҳиссаҳои номустақил (ёрирасон)-и нутқанд. Ду ҳиссаи нутқи охиринро (калимаи тақлидиовозӣ ва нумератив) ба ҳиссаҳои нутқи шаклсоз дохил мекунем. Панҷ ҳиссаи нутқи дигар категорияҳои шаклсоз надоранд<sup>2</sup>. Дар луғат ва матнҳо калимаҳои алоҳидаи ба ин ҳиссаҳои нутқ дохилшаванда, ба ғайр аз ҳолатҳои истисноӣ, дар намуди тағйирнопазир вомехӯранд. Аз ин сабаб, таҳлили морфологии онҳо сода буда, танҳо идентификатсияи элементҳои мувофиқро дар базаи маълумот талаб мекунад. Бинобар ин, дар ин боб танҳо категорияҳои шаклсози ёздаҳ ҳиссаҳои нутқи қайдшуда матраҳ мегарданд.

Дар ҳамаи параграфҳои минбаъдаи боби мазкур, ҳангоми тавсифи категорияҳои грамматикӣ ҳиссаҳои нутқ, тартиби пайвастшавии аффиксҳои шаклсозро ба *формаи нормалӣ* баён мекунем. Ин пайвастшавӣ боиси ҳосил шудани калимаҳои мегардад, ки *калимаҳои шаклсоз* номида мешаванд. ***Дар навбати худ, бо назардошти хусусиятҳои забони тоҷикӣ, ба сифати формаи нормалӣ ё реша ва ё асоси калима истифода мешавад.*** Аз ин ду мафҳуми асос нисбатан маъмул мебошад, аз ин хотир, он ба сифати формаи нормалӣ калима истифода хоҳад шуд.

Дар ин боб кодиронии мавқеи калимаҳои ба ҳиссаҳои гуногуни нутқ тааллуқдошта муайян карда мешавад. Ду мавқеи аввали коди калима барои муайян намудани ҳиссаи нутқи калима чудо шудааст, барои исм – 01, сифат – 02, шумора – 03, ҷонишин – 04, феъл – 05, масдар – 06, сифати феълӣ – 07, феъли ҳол – 08, зарф – 09, пешоянд – 10, пасоянд – 11, пайвандак – 12, ҳиссача – 13, нидо – 14, калимаи тақлидиовозӣ – 15, нумератив – 16. Ҳамин тавр, аз рӯи ду рақами коди калима мансубияти он ба ҳиссаи нутқ муайян карда мешавад. Рақамҳои боқимонда ин коди қиматҳои нишонаҳои грамматикӣ (граммемаҳо) мебошанд.

Дар § 2.1 барои кодиронии калимаҳои ба исм тааллуқдошта чунин қолаб пешниҳод шудааст:

01	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$
----	------------	------------	------------	------------

Ба символҳои  $\alpha_k$  ( $k = \overline{1,4}$ ) коди 0 ва ё 1 бахшида мешавад, ки ба онҳо граммемаҳои муайян ва постфиксҳои онҳоро амаликунанда аз рӯи қоидаи зерин мувофиқ меоянд:

<sup>2</sup> Ба истисноӣ пешоянди барои, ки дар матн дар намудҳои бароям, бароят ва ғ. мушоҳида мешавад.

$$\alpha_1 = \begin{cases} 0 - \text{шумораи танҳо } (\emptyset); \\ 1 - \text{шумораи чамъ} \\ (\text{вон, гон, он(ён), хо, чот}). \end{cases} \quad \alpha_3 = \begin{cases} 0 - \text{набудани бандаки изофи } (\emptyset); \\ 1 - \text{мавчудияти бандаки изофат}(u). \end{cases}$$

$$\alpha_2 = \begin{cases} 0 - \text{муайяни } (\emptyset); \\ 1 - \text{номуайяни } (e). \end{cases} \quad \alpha_4 = \begin{cases} 0 - \text{набудани пасоянд } (\emptyset); \\ 1 - \text{мавчудияти пасоянд } (po). \end{cases}$$

Дар ин қоидаҳо симболи  $\emptyset$  ифодакунандаи постфикси холӣ (мавчуд набудани постфикс) мебошад. Дар поён мисоли кодиронии калимаҳои ба исм тааллуқдошта оварда шудааст:

*Мисоли 1.* китобҳоеро = китоб  $\oplus$  хо  $\oplus$  е  $\oplus$   $\emptyset$   $\oplus$  ро = 011101.

*Мисоли 2.* китобҳои = китоб  $\oplus$  хо  $\oplus$   $\emptyset$   $\oplus$  и  $\oplus$   $\emptyset$  = 011010.

*Мисоли 3.* китоб = китоб  $\oplus$   $\emptyset$   $\oplus$   $\emptyset$   $\oplus$   $\emptyset$   $\oplus$   $\emptyset$  = 010000.

Дар ин мисолҳо аввал, калима оварда шуда, сипас, калима ба морфҳо чудо мегардад ва дар охир коди он оварда мешавад. Символи  $\oplus$  пайвастшавиро ифода мекунад.

Дар §§ 2.2 – 2.11 чунин қолабҳо барои кодиронии калимаҳои ба дигар ҳиссаҳои нутқ тааллуқдошта (сифат, шумора, ҷонишин, феъл бо асоси замони ҳозира ва гузашта, масдар, сифати феълӣ, феъли ҳол, зарф, калимаи тақлидиовозӣ ва нумератив) пешниҳод карда шуда, намунаи кодиронии калимаҳо оварда мешавад.

Дар боби 1 дар натиҷаи татбиқи шиносандаи морф ба коркарди нимаавтоматии маҷмӯи матнҳо 79702 формаҳои нормалии калимаҳои тоҷикӣ ба даст омадааст, ки аз онҳо 50984 реша ва 28718 асоси калима мебошанд. Дар боби 2 моделҳои математикии ташкили калимаҳо аз 16 ҳиссаи нутқ тавсиф шудааст. Моделҳо ба намуди кодиронии мавқеи калимаҳо ифода гардидаанд. Кодиронии мавқеи имконият медиҳад, ки бо роҳи ба кодҳо бахшидани қиматҳои имконпазир (ба қиматҳои конкретии грамматикӣ категорияҳои шаклсоз мувофиқ меоянд) шаклҳои гуногуни парадигмаҳои формаҳои нормалии ба ҳиссаҳои гуногуни нутқ тааллуқдошта ҳосил карда шаванд.

Дар **боби 3** мо ба базаи бузурги аффиксҳои забони тоҷикӣ, ки дар калимаҳо дар намуди ҳамчояшавии намоёндаҳои элементарии онҳо вохӯранд, муроҷиат мекунем. Бо назардошти хусусиятҳои калимасозӣ дар забони тоҷикӣ ба кластеризатсияи қабулшудаи аффиксҳо - *шаклсоз* ва *калимасоз* мо намуди иловагии *иборасозро* ҳамроҳ мекунем. Мундариҷаи асосии боб тавсифи базаи маълумот оид ба тағйирёбии ҳиссаи нутқи калима дар натиҷаи пайваст шудани аффиксҳои сода мебошад. Натиҷаи омӯзиш ва чамъоварии пайвастшавии аффиксҳои сода ба ҳаммаи ҳиссаҳои нутқ заминаи татбиқи таҳлили автоматии морфологӣ калимаҳоро ташкил доданд.

Ифодаи калимаи дилхоҳро ба намуди  $WF = PR \oplus R \oplus PS$ , ки пештар истифода шуда буд, ҳамчун инъикоси

$$R \xrightarrow{f} WF \quad (3.1)$$

мачмӯи решаҳо  $R = \{R\}$  ба мачмӯи  $WF = \{WF\}$  калимаҳои забони тоҷикӣ, ки бо воситаи  $f$  – оператор амалӣ мешавад, тасвир кардан мумкин аст.  $f$  – оператор бо чунин намудҳо истифода мешавад:

оператори мушобех  $f = E$ , агар  $WF = R$ ;

префикс-оператор  $f = PR$ , агар  $WF = PR \oplus R$ ;

постфикс-оператор  $f = PS$ , агар  $WF = R \oplus PS$ ;

префикс  $\wedge$  постфикс-оператор  $f = PR(\cdot)PS$ , агар  $WF = PR \oplus R \oplus PS$ .

Дар шарҳи мазкур  $WF$  ҳамчун натиҷаи тартибдиҳии калима бо роҳи пайвастиши аффиксҳои шаклсоз, калимасоз ва иборасоз нишон дода мешавад. Дар ин ҳолат калимаи  $WF$  –ро ҳамчун *образи* решаи  $R$  ва решаи  $R$  –ро ҳамчун *образи аввалияи* калимаи  $WF$  меҳисобанд.

Новобаста ба он ки ғояи тартибдиҳии калима бо мақсади тавсифи хосиятҳои он комилан фаҳмо аст, татбиқи ин ғоя дар амал кори сахл нест. Сабаби асосӣ структураи мураккаби аффиксҳои пайвастишаванда – префиксҳо (то 3 дараҷа) ва постфиксҳо (то 8 дараҷа) аст. Аз ин сабаб, ташхиси калимаи дорои миқдори муайяни аффиксҳо марҳила ба марҳила амалӣ карда мешавад. Ташхис аз реша, ки ба он аффикси наздиктарин пайваст мешавад, оғоз мегардад, сипас, ба фрагменти ҳосилшуда боз як аффикси дигар пайваст мешавад, ин амал то аффикси содаи охири идома меёбад. Дар ҳар як итератсия ҳиссаи нутқ ва маъноӣи фрагмент муайян карда мешавад. Ин амал имкон медиҳад, ки дар умум категорияҳои грамматикӣ ва граммемаҳои калима муайян шавад.

**Намудҳои калима (§ 3.1).** Муайян намудани ҳиссаи нутқи баъзе калимаҳои забони тоҷикӣ муяссар намешавад. Масалан, мафҳуми “бӯямат” ба ягон ҳиссаи нутқ марбут нест, зеро он аз ҷиҳати мазмун ба иборати “ман туро бӯям” ҳаммаъно аст. Чунин намуди калимаҳоро калимаҳои *иборасоз* меномем.

Ташхиси омории мачмӯи матнҳои иборат аз 59344883 калима муайян намуд, ки калимаҳои иборасоз 27.25 % -и калимаҳои уникалӣ ва 4.78% -и ҳамаи калимаҳоро ташкил медиҳанд. Дар асоси ин маълумот зарурати ба инобат гирифтани намуди сеюми калимаҳо ба миён меояд.

**Таърифи 3.3.** *Калимаҳои иборасоз – натиҷаи инъикос (3.1), ки дар он образ (калимаи ҳосилшаванда) аз ҷиҳати семантикӣ ба ибора (фрагменти ҷумла) ҳаммаъно аст.*

Мисолҳо: мафҳуми «*дидамат*» ба иборати «*ман туро дидам*»; «*гуфтамаиш*» - ба иборати «*ман ба ӯ гуфтам*» ҳаммаъно аст.

Дар § 3.2 системаи аломатҳои барои шарҳи натиҷаи татбиқи аффикс-операторҳо ба ҳиссаҳои мустақили нутқ истифодашуда фаҳмонида мешавад. Дар он барои се намуди калимаҳо ва аффиксҳои мувофиқи онҳо код (1-шаклсоз, 2-калимасоз, 3-иборасоз) дохил карда мешавад.

Дар § 3.3 натиҷаи пайвастиши аффиксҳо ба образи аввалияе, ки ба исми таллуқ дорад, оварда мешавад (нигаред ба чадвали 3.4).

Чадвали 3.4. - Пайвастшавии аффиксҳо ба исм

№	Префикс	Миқдор	Коди ҳиссаи нутқ	Коди намуди калима
a	ҳам-, бар-,	2	01	2
b	ба-, бе-, беҳам-, бо-, но-, ноҳам-, ноба-, то-, бар-, баҳам-, ноҳам-	11	02	2
c	бано-,	1	09	2
d	би-, ма-, ме-, на-, наме-, ҳама-, бу-, мена-, наби-, меби-, ҳамена-, ҳамеби-, ноби-,	13	05	2
	Постфикс			
e	+й, +й-ву, +й-ю, +ам, +ям, +ам-й, +ям-й, +ам-й-ву...	166	8	3
f	+и, +ро,	2	01	1
g	бон, ... гар, ... гор,.. истон,..манд,..навард,..сор...,	1389	01	2
h	ваш, ... ворона, ... гин, гун, ... ин,.. манд, ... монанд, ...нок...,	88	02	2
i	ид, онд, ёнд, онид, ёнид	5	05	2
j	идан, ондан, ёндан, онидан, ёнидан	5	06	2
k	анда, янда, ида, идагӣ, онда, ёнда, ондагӣ, ёндагӣ, онида, ёнида, онидагӣ, ёнидагӣ, иданӣ,	13	07	2
l	ан, ян, гарона, мандона, мандонатар, сон, сонӣ,ангӣ, янгӣ	9	09	2
m	ӣ, акӣ, бор, борӣ, вона, вор, ворӣ, гӣ, гинӣ, гоҳӣ, инӣ, ина, инча, манд, мандӣ, ноки, ой, онӣ, отӣ, сор, сорӣ, якӣ	22	01/02	2
n	истонӣ, нгӣ, стонӣ, шанӣ	4	01/09	2
o	вон, ... гон, ... е, ... он, ... ён, ... ҳак, ...ҷот,...	57	01	1

Дар чадвал 5 сутун мавҷуд аст. Дар сутуни 1-ум бо истифодаи ҳарфҳои латинӣ гурӯҳи аффиксҳо (префиксҳо ва ё постфиксҳо) ишора шудаанд. Элементҳои ин гурӯҳҳо дар сутуни дувум оварда шудаанд. Дар сутуни сеюм миқдори аффиксҳои гурӯҳи муоинашаванда нишон дода шудааст. Дар сутуни 4-ум ҳиссаи нутқи калимае, ки пас аз пайваст шудани аффикси дилхоҳи гурӯҳ ҳосил мешавад, оварда шудааст. Дар сутуни 5-ум намуди калимаи ҳосилшуда нишон дода мешавад (намуди аффикси пайвастшуда).

Дар замимаҳои 3.1-3.8. натиҷаи пайвастшавии аффиксҳо ба ҳиссаҳои мустақили нутқ, калимаҳои тақлидиовозӣ ва нумератив оварда шудааст.

**Боби 4** ба алгоритмиронии раванди таҳлили морфологӣ калимаҳои забони тоҷикӣ, ки аз 4 марҳила иборат аст, бахшида шудааст:

1. *Ҷудо намудани калима ба морфҳо, ки дар чунин намуд ифода мешавад:*

$$WF = PR \oplus R \oplus PS^1 \oplus PS^2, \quad (4.1)$$

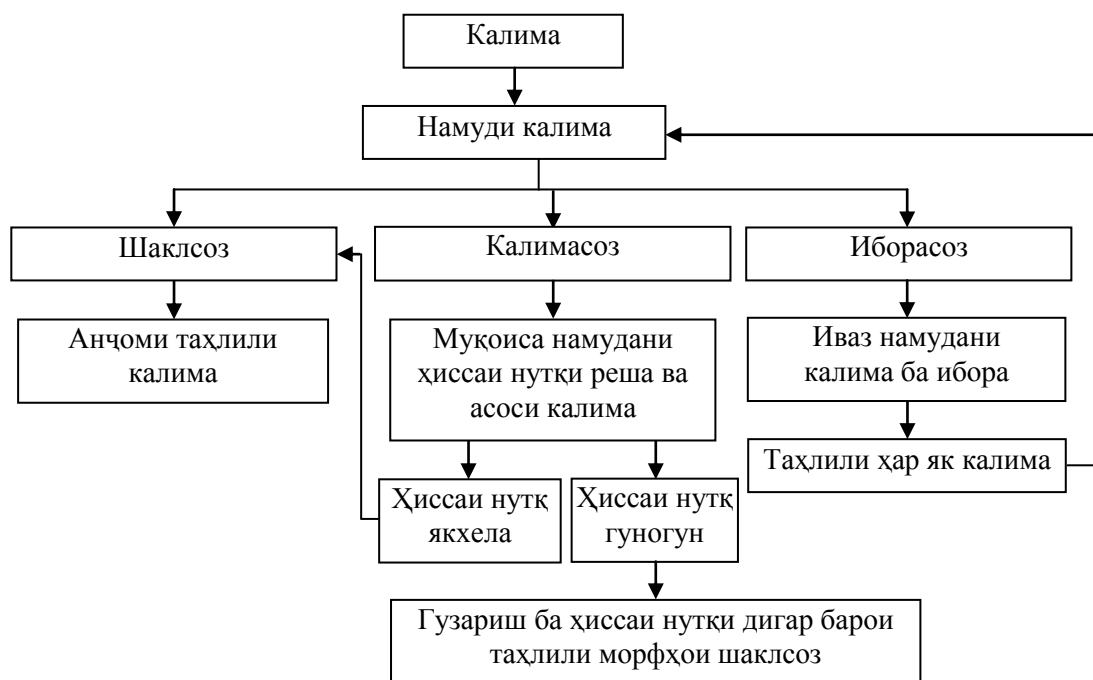
ки ин ҷо  $R$  – реша ва  $PR$  – префикс. Суффикси  $PS^1$  ва банди  $PS^2$  ҳамчун постфиксҳои парадигмаофар ва парадигмасоз шинохта мешаванд;

2. Муайян намудани намуди калимаи додашуда;
3. Муайян намудани ҳиссаи нутқ ва граммаҳои реша  $R$ ;
- 4.1. Муайян намудани асос, ҳиссаи нутқ ва граммаҳои калима, агар калима

ба намуди шаклсоз ва ё калимасоз тааллуқ дошта бошад;

4.2. Пешниҳоди калима ба намуди ибора ва муайян намудани асос, ҳиссаи нутқ ва граммемаҳои ҳар як калимаи таркиби ибора, агар калима ба намуди иборасоз тааллуқ дошта бошад.

Шиносандаи морф, ки дар боби 1 сохта шудааст, имконияти иҷрои ду процедураи аввали таҳлилгари морфологиро дорад. Қисмати асосии системаи таҳлили морфологиро алгоритми коркарди калимаҳои аз решаҳои ин ё он ҳиссаи нутқ ҳосилшуда ташкил медиҳанд. Ин алгоритмҳо (новобаста аз мақсади гуногун) тартиби ягонаи таҳлили калимаҳои намуди гуногунро ифода мекунанд, ки моҳияти он дар расми 4.1 оварда мешавад.



Расми 4.1. - Тартиби коркарди калимаҳо

Монандии алгоритмҳо дар коркарди намуди гуногуни калима зоҳир мегардад. Агар намуди  $WF$  шаклсоз бошад, он гоҳ таҳлили он дар доираи алгоритми коркарди калимаи ин ҳиссаи нутқ бапурагӣ ба анҷом мерасад.

Агар намуди калима иборасоз бошад, он гоҳ калима бар ибора иваз шуда, таҳлили морфологии ҳар як калима дар алоҳидагӣ иҷро мешавад.

Агар намуди калима калимасоз бошад, он гоҳ таҳлили асоси он, ки намуди умумии  $PR \oplus R \oplus PS^1$ -ро дорад, иҷро карда мешавад. Вобаста ба натиҷа ду ҳолат имконпазир аст:

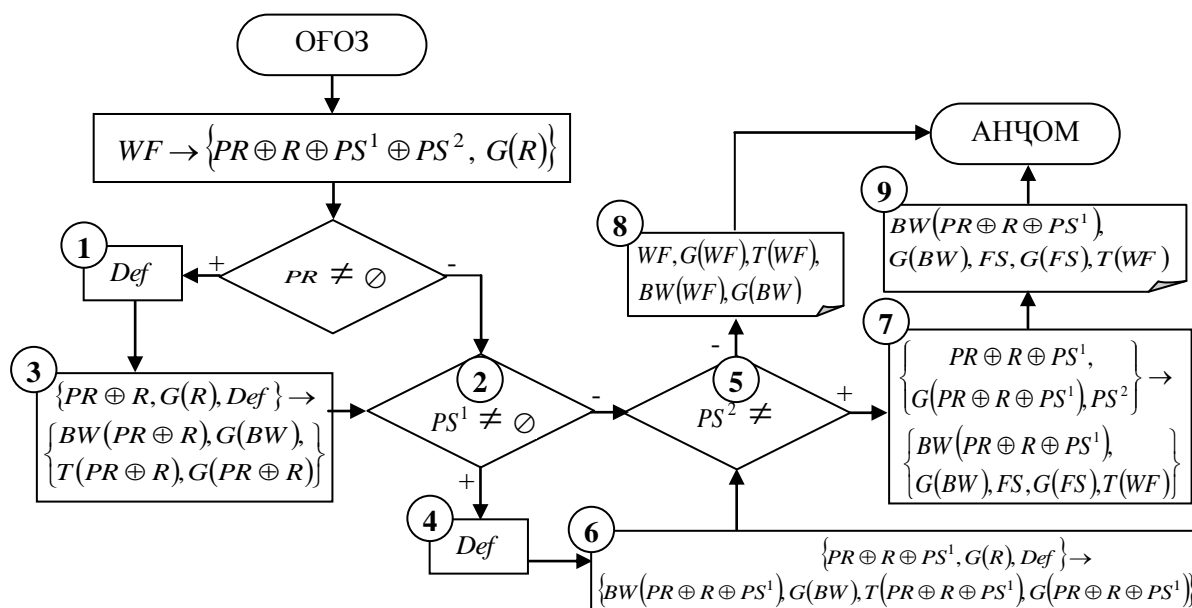
– реша  $R$  ва асоси он  $PR \oplus R \oplus PS^1$  ба як ҳиссаи нутқ тааллуқдоранд; дар ин ҳолат намуди калима нисбат ба асоси он калимаи шаклсоз ҳисоб гардида, таҳлили он бапурагӣ ба анҷом мерасад;

– реша  $R$  ва асоси он  $PR \oplus R \oplus PS^1$  ба ҳиссаҳои гуногуни нутқ тааллуқдоранд; дар ин ҳолат алгоритми таҳлили калима бо решае, ки ба ҳиссаи нутқи муа-



йян тааллуқ дорад, кори худро ба анҷом мерасонад. Қисмати боқимондаи таҳлил аз тарафи алгоритми таҳлили калимаҳое, ки решаи онҳо ба ҳиссаи нутқе, ки ба ҳиссаи нутқи асоси  $PR \oplus R \oplus PS^1$  калимаи омӯхташаванда мувофиқат мекунад, иҷро карда мешавад.

Барои таҳлили морфологии калимаҳо таҳлилгари компютери морфологӣ, ки аз шиносандаи морфҳо ва 16 зерсистема иборат аст, тартиб дода шудааст. Модели чамъбасти таҳлилгари морфологӣ дар расми 4.2. оварда шудааст.



Расми 4.2. - Модели чамъбасти таҳлили морфологии калимаҳо

Хусусияти чамъбасти модел дар он ифода мегардад, ки дар модел, аз як тараф, концепсияи амалкунии ҳамаи 16 зерсистемаҳо ва аз тарафи дигар, концепсияи амалкунии система пурра дар бар гирифта мешавад.

Дар ин расм бар иловаи  $WF$ ,  $PR$ ,  $R$ ,  $PS^1$  ва  $PS^2$  ишораҳои зерин истифода мешаванд:  $BW$  – асоси калима,  $FS$  – калимаи иборасоз, ки аз рӯи мазмун ба ибора баробарарзиш аст (фрагменти ҷумла);  $G(f)$ ,  $BW(f)$ ,  $T(f)$  - функцияҳо барои муайян намудани граммемаҳо, асос ва намуди (шаклсоз, калимасоз ва иборасоз) аргументи  $f$ , ки қиматҳои  $PR \oplus R$ ,  $R$ ,  $PR \oplus R \oplus PS^1$  ва  $R \oplus PS^1$  -ро қабул карда метавонад. Илова бар ин, намуди префикс (постфикс) бо  $Def$  ишора шудааст. Агар намуди морф шаклсоз бошад, он гоҳ  $Def = 1$ , дар акси ҳол  $Def = 0$ .

Мувофиқи расми 4.2 таҳлили морфологӣ аз дохил кардани калима  $WF$  оғоз мегардад. Шиносандаи морфи дар боби 1 овардашуда ду протокураи вазифаи таҳлили морфологиро иҷро мекунад:

- $WF$  –ро ба морфҳо ҷудо мекунад,
- ҳиссаи нутқ ва рӯйхати граммемаҳои реша  $R$  –ро муайян мекунад.

Сипас, мавҷудияти префикс дар таркиби калимаи таҳлилшаванда тафтиш карда мешавад. Агар мавҷуд бошад, амалиёти қисми 1, дар акси ҳол, амалиёти

қисми 2 ичро мешавад.

Дар қисми 1 вобаста аз намуди префикс (шаклсоз ё не) ба симболи  $Def$  қимати 1 ё 0 бахшида мешавад. Сипас, гузариш ба қисмати 3 амалӣ мешавад. Дар қисмати 3 дар асоси фрагменти  $PR \oplus R$ -и калимаи  $WF$ , граммемаҳои реша  $R$  ва маълумоти  $Def$  оид ба намуди префикс асоси  $BW(PR \oplus R)$  фрагменти  $PR \oplus R$ , граммемаҳои асос  $G(BW)$ , намуд  $T(PR \oplus R)$  ва граммемаҳои  $G(PR \oplus R)$ -и фрагменти  $PR \oplus R$  муайян карда мешавад.

Қайд кардан зарур аст, ки фрагменти  $PR \oplus R$  метавонад, бо асос мувофиқ наояд. Агар намуди префикс  $PR$  шаклсоз бошад, он гоҳ  $BW(PR \oplus R) = R$  мешавад. Масалан, дар калимаи “*мерафт*” намуди префикси  $PR = “me”$  шаклсоз аст. Бинобар ин,  $BW(me \oplus рафт) = рафт$ .

Агар намуди префикс  $PR$  калимасоз бошад, он гоҳ  $BW(PR \oplus R) = PR \oplus R$  мешавад. Масалан, дар калимаи “*боодоб*” намуди префикси  $PR = “бо”$  калимасоз аст. Бинобар ин,  $BW(бо \oplus одоб) = “боодоб”$ , яъне асос ва фрагмент ба ҳам баробаранд.

Дар қисми 2 вобаста аз мавҷудияти постфикси  $PS^1$  гузариш ба қисми 4 (мавҷуд будан) ва ё қисми 5 (мавҷуд набудан) амалӣ мешавад.

Дар қисми 4 вобаста аз намуди постфикс (шаклсоз ё не) ба симболи  $Def$  қимати 1 ё 0 бахшида мешавад. Пас аз он гузариш ба қисмати 6 амалӣ мешавад. Дар қисмати 6 дар асоси фрагменти  $PR \oplus R \oplus PS^1$ -и калимаи  $WF$ , граммемаҳои реша  $R$  ва маълумоти  $Def$  оид ба намуди постфикс, асоси  $BW(PR \oplus R \oplus PS^1)$  фрагменти  $PR \oplus R \oplus PS^1$ , граммемаҳои асос  $G(BW)$ , намуд  $T(PR \oplus R \oplus PS^1)$  ва граммемаҳои  $G(PR \oplus R \oplus PS^1)$ -и фрагменти  $PR \oplus R \oplus PS^1$  муайян карда мешаванд.

Қайд кардан зарур аст, ки фрагменти  $PR \oplus R \oplus PS^1$  метавонад бо асос мувофиқ наояд. Агар намуди постфикси  $PS^1$  шаклсоз бошад, он гоҳ  $BW(PR \oplus R \oplus PS^1) = (PR \oplus R) \wedge R$  (симболи  $\wedge$  ифодакунандаи “ё” –и мантиқӣ аст) мешавад. Масалан, дар калимаи “*ҳампешагон*” намуди постфикси  $PS^1 = «гон»$  шаклсоз аст. Бинобар ин,  $BW(хам \oplus пеша \oplus гон) = хампеша$ .

Агар намуди постфикси  $PS^1$  калимасоз бошад, он гоҳ  $BW(PR \oplus R \oplus PS^1) = (PR \oplus R \oplus PS^1) \wedge R \oplus PS^1$  мешавад. Масалан, дар калимаи “*ҳампешагӣ*” намуди постфикси  $PS^1 = “гӣ”$  калимасоз аст. Бинобар ин,  $BW(хам \oplus пеша \oplus гӣ) = хампешагӣ$ , яъне, асос ва фрагмент ба ҳам баробаранд.

Дар қисми 5 мавҷудияти  $PS^2$  тафтиш карда мешавад. Агар мавҷуд бошад, гузариш ба қисми 7 дар акси ҳол, ба қисми 8 амалӣ мегардад.

Дар қисмати 8 дар асоси фрагменти  $PR \oplus R \oplus PS^1$ -и калимаи  $WF$ , граммемаҳои  $G(PR \oplus R \oplus PS^1)$ , фрагменти  $PR \oplus R \oplus PS^1$  ва постфикси  $PS^2$  асоси  $BW(PR \oplus R \oplus PS^1)$  фрагменти  $PR \oplus R \oplus PS^1$ , граммемаҳои асос  $G(BW)$ , намуд  $T(WF)$  ва граммемаҳои  $G(PR \oplus R \oplus PS^1)$ -и фрагменти  $PR \oplus R \oplus PS^1$  муайян карда мешавад. Сипас, бо истифодаи алгоритми махсус, ки дар [9] пешниҳод шуда-

аст, калимаи иборасоз ба намуди ибора  $FS$  в граммемаҳои ҳар як калимаи ибора  $G(FS)$  ифода мешавад.

Дар қисмҳои 8 ва 9 натиҷаи ниҳонии таҳлили морфологии калима оварда мешавад.

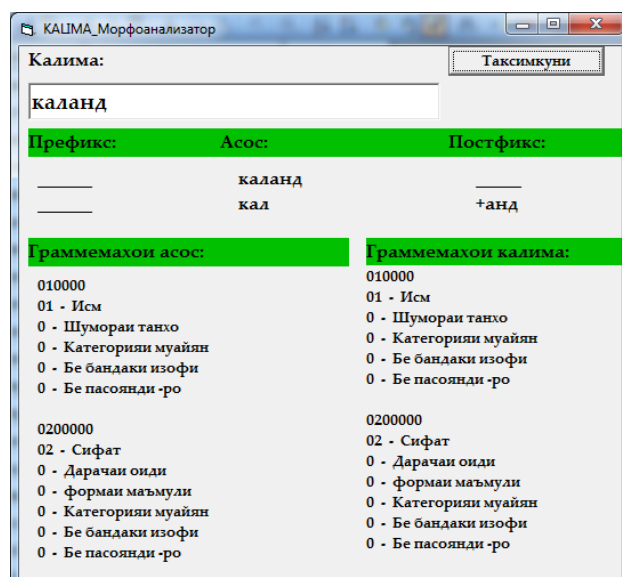
Дар §§ 4.2-4.11 алгоритми таҳлили морфологии калимаҳо, ки решаашон ба ҳиссаҳои мустақили нутқ тааллуқ доранд, оварда шудааст. Дар § 4.12 алгоритми табдили калимаҳои иборасоз ба ибораҳо (фрагменти ҷумла) оварда шудааст.

Дар **боби 5** тавсифи комплекси барномавӣ, ки 4 марҳилаи таҳлили морфологиро иҷро мекунад, оварда шудааст. Намуди ҷамъбастии комплекси барномаҳои таҳлили морфологӣ дар расми 5.1 нишон дода шудааст.



Расми 5.1. - Сохтори схемавии барномаи комплексӣ

Натиҷаи таҳлили морфологии калима ба интерфейси барнома бароварда мешавад, нигаред ба расми 5.2.



### Расми 5.2. - Интерфейси барномаи комплексӣ

Дар § 5.4 сандиши самаранокии барномаи комплексӣ дар асоси коркарди интихоби тасодуфӣ – файли матнии иборат аз 314134 калима гузаронида шуд. Аз 41824 калимаҳои уникалии ба барномаи комплексӣ пешниҳодшуда 38487 калима, яъне, 92,02 % калимаҳо таҳлил шуданд.

Ҳамин тавр, таҳлилгари морфологӣ бо самаранокии 92,02 %-ро, ки дар доираи диссертатсия сохта шудааст, барои истифодаи назариявӣ ва амалӣ қобили қабул ҳисобидан мумкин аст.

## ХУЛОСА

Натиҷаҳои асосии диссертатсия чунинанд:

1. Тартиб додани шиносандаи морф – ҳамчоя намудани протсекураи итеративии нимаавтоматии компютери тартиб додани базаи морфҳо бо роҳи коркарди файлҳои матнӣ ва “ташҳиси дастӣ” –и ҷудокунии калимаҳои коркарднашуда.

2. Тартиб додани базаи морфҳои забони тоҷикӣ аз 82 префикс, 76 539 реша ва 128 760 постфикс (бо роҳи ташкил намудани маҷмӯи матнҳои иборат аз 59 344 883 калима ва коркарди он бо татбиқи шиносандаи морф). Бо истифодаи методҳои омӯрӣ тасдиқ намудани мукаммали базаи префиксҳо, эҳтимолияти дар оянда нисбатан зиёд гаштани шумораи постфиксҳо ва бемаҳдудият зиёд гаштани рӯихати решаҳо.

3. Пешниҳод намудани таснифи “шаклсоз-калимасоз-иборасоз” барои аф-фиксҳо ва калимаҳои забони тоҷикӣ.

4. Коркарди кодиронии маҷеии калимаҳои забони тоҷикии ба ҳиссаҳои гуногуни нутқ тааллуқдошта дар асоси тавсифи системавии категорияҳои шаклсоз, граммемаҳои реша ва асоси калимаҳо.

5. Ба шакл даровардани қоидаҳои калимасозии забони тоҷикӣ;

6. Коркарди комплекси барномаҳои таҳлили автоматии морфологии калимаҳои забони тоҷикӣ.

Натиҷаҳои диссертатсия бо ҳамчоягии методикаи Hunspell ҳамчун заминаи тартиб додани пакетҳои забонӣ барои тафтиши имлои калимаҳои забони тоҷикӣ дар OpenOfficeOrg ва Microsoft Office истифода шуд. Пакетҳои мазкур ҳамчун захираҳои иттилоотии ЗИ-03.2.222ТJ аз 11.01.2012 ва № 4201200235 аз 04.10.2012 ба қайд гирифта шудаанд. Онҳо дар аксари муассисаҳо ва корхонаҳои Ҷумҳурии Тоҷикистон истифода мешаванд.

Системаи компютери таҳлили морфологӣ заминаи асоси коркарди барномаҳои гуногуни коркарди автоматии матнро ташкил медиҳад. Ҳангоми коркарди барномаҳои коркарди автоматии матн қисмати зиёди протсекураҳои техникиро системаи компютери таҳлили морфологӣ иҷро мекунад.

## РҶҶҲАТИ ИНТИШОРОТ ДАР МАВЗҶИ ДИССЕРТАТСИЯ

1. Довудов, Г.М. О формировании базы префиксов таджикского литературного языка [Текст] / Г.М. Довудов, З.Д. Усманов // Доклады АН РТ. - 2009. - Том 52, №6. - С. 431-436.
2. Довудов, Г.М. О множестве постфиксов таджикского литературного языка [Текст] / Г.М. Довудов, З.Д. Усманов, О.М. Солиев // Доклады АН РТ. - 2010. - Том 53, №2. - С. 99-103.
3. Довудов, Г.М. О статистических закономерностях морфемной базы таджикского языка [Текст] / Г.М. Довудов, З.Д. Усманов // Доклады АН РТ. - 2010. - Том 53, №3. - С. 188-191.
4. Довудов, Г.М. Частотный морфемный словарь таджикского литературного языка [Текст] / Г.М. Довудов, З.Д. Усманов // Доклады АН РТ. - 2010. - Том 53, №4. - С. 257-262.
5. Довудов, Г.М. Концептуальная модель автоматического морфологического анализа таджикских словоформ [Текст] / Г.М. Довудов, З.Д. Усманов // Доклады АН РТ. - 2014. - Том 57, №3. - С. 205-209.
6. Довудов Г.М. Алгоритм автоматического морфологического анализа таджикских слов [Текст] / Г.М. Довудов // Известия АН РТ. - 2010. - №2 (139). - С. 22-26.
7. Довудов, Г.М. Статистика частей речи таджикского языка [Текст] / Г.М. Довудов // Известия АН РТ. - 2012. - №3 (148). - С.54-56.
8. Довудов, Г.М. Алгоритм представления таджикских словосочетательных словоформ фрагментами предложений [Текст] / Г.М. Довудов, З.Д. Усманов // Известия АН РТ. - 2013. - №4 (153). - С. 69-75.
9. Довудов, Г.М. Позиционное кодирование таджикских словоформ [Текст] / Г.М. Довудов, З.Д. Усманов // Известия АН РТ. - 2015. - №1(158). - С.58-66.
10. Довудов, Г.М. Формирование базы морфов таджикского языка [Текст]: монография / Г.М. Довудов, З.Д. Усманов. - Душанбе: Дониш, 2014. - 109 с.
11. Довудов, Г.М. Морфологический анализ словоформ таджикского языка [Текст]: монография/Г.М. Довудов, З.Д. Усманов. - Душанбе: Дониш, 2015. - 132 с.

**Дар нашриёти илмии дигар** 4 шаходатнома дар бораи бақайдгирии давлатии захираи иттилоотӣ дар Маркази миллии патенту иттилоотии Вазорати рушди иқтисодиёт ва савдои Ҷумҳурии Тоҷикистон ва 4 мақола дар маҷмӯаҳои Masaryk University ва European Language Resources Association (ELRA) интишор гардидааст.

## ШАРҲИ МУХТАСАР

**ба рисолаи диссертатсионии Довудов Гулшан Мирбахоевич дар мавзӯи  
«Таҳлили морфологии компютерии калимаҳои забони тоҷикӣ»  
барои дарёфти дараҷаи илмии номзоди илмҳои техникӣ аз рӯи ихтисоси  
05.13.11 - Таъминоти математикӣ ва барномавии мошинҳои ҳисоббарор,  
мучтамаъҳо ва шабакаҳои компютерӣ**

**Калимаҳои калидӣ:** забони тоҷикӣ, таҳлили автоматии морфологӣ.

**Объекти таҳқиқот:** қонуниятҳои грамматикии забони тоҷикӣ.

**Мақсади кор:** Алгоритмиронии раванди таҳлили морфологии калимаҳои забони тоҷикӣ ва татбиқи он дар намуди комплексӣ барномавӣ.

**Методҳои таҳқиқот:** назарияи маҷмӯъҳо, таҳлили системавӣ, элементҳои назарияи кодиронӣ, моделиронии математикӣ, омори математикӣ, барномарезии баобъектнигаронидашуда.

**Натиҷаҳои бадастомада ва нағсонии онҳо:** Натиҷаҳои асосии диссертатсия нағб буда, чунин ифода мегарданд:

- базаи бузурги морфҳои забони тоҷикӣ ташкил карда шуд, ки аз 81 префикс, 76 539 реша ва 128 760 постфикс иборат аст;

- таснифи намуди аффиксҳо (шаклсоз, калимасоз ва иборасоз) ва калимаҳо пешниҳод шуд;

- кодиронии мавқеии калимаҳои забони тоҷикӣ коркард шуд;

- алгоритми табдили калимаҳои иборасоз ба ибораҳо пешниҳод шуд;

- таъминоти алгоритмии таҳлили морфологии калимаҳои забони тоҷикӣ коркард шуд.

**Аҳмияти амалии таҳқиқот:** дар асоси таҳлилгари морфологӣ пакетҳои забони барои тафтиши имлоии калимаҳои забони тоҷикӣ дар OpenOfficeOrg ва Microsoft Office тартиб дода шудааст, ки ҳамчун захираҳои иттилоотии ЗИ-03.2.222ТJ аз 11.01.2012 ва № 4201200235 аз 04.10.2012 ба қайд гирифта шудаанд. Пакетҳои мазкур дар аксари муассисаҳо ва корхонаҳои Ҷумҳурии Тоҷикистон истифода мешаванд.

**Соҳаи истифодабарӣ:** ҳалли масъалаҳои коркарди автоматии матн, тарҷумаи компютерӣ, тафтиши имло, таҳлил ва синтези нутқ, муколама бо компютере, автоматизатсияи равандҳои индексиронӣ, шарҳдиҳӣ, гуруҳбандӣ, фишурданамоӣ, ҷудо намудани калимаҳои калидӣ, ва тафсили матн.

## АННОТАЦИЯ

диссертации Довудова Гулшана Мирбахоевича на тему «Компьютерный морфологический анализ таджикских словоформ»  
на соискание ученой степени кандидата технических наук по специальности  
05.13.11 – Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

**Ключевые слова:** таджикский язык, автоматический морфологический анализ.

**Объект исследования:** грамматические закономерности таджикского языка.

**Цель работы:** Алгоритмизировать процесс морфологического анализа таджикских словоформ и реализовать его в виде программного комплекса.

**Методы исследования:** теория множеств, системный анализ, элементы теории кодирования, математическое моделирование, математическая статистики, объектно-ориентированное программирование.

**Полученные результаты и их новизна:** Результаты диссертации являются новыми и заключаются в следующем:

- сформирована обширная база морфов таджикского языка, содержащая 81 префикс, 76 539 корней и 128 760 постфиксов;
- предложена классификация типов аффиксов (словоизменяемых, словообразовательных и словосочетательных) и аналогичная классификация словоформ;
- разработано позиционное кодирование таджикских словоформ;
- разработано эквивалентное представление словосочетательных словоформ фрагментами предложения;
- разработано алгоритмическое обеспечение морфологического анализа таджикских словоформ.

**Степень использования:** на основе морфологического анализатора созданы языковые пакеты для проверки таджикской орфографии в OpenOfficeOrg и Microsoft Office, зарегистрированные в виде информационных ресурсов ЗИ-03.2.222ТJ от 11.01.2012 г. и № 4201200235 от 04.10.2012 г. Пакеты используются во многих организациях и учреждениях РТ.

**Область применения:** решение задач автоматической обработки текста, компьютерный перевод, проверка орфографии, анализ и синтез речи, диалог с компьютером, автоматизация процессов индексирования, аннотирования, реферирования, классификации и рубрикации документов, извлечения ключевых слов, морфологической разметки корпуса таджикского языка.



## ANNOTATION

**on the dissertation of Dovudov Gulshan Mirbahoevich on the theme «Computer morphological analysis of Tajik word form (slovoform)» candidate for a degree of technical sciences on a specialty**

**05.13.11 – Mathematical and software of computers, complexes and computer networks**

**Key words:** Tajik, automatically morphological analysis.

**Research object:** Grammatical rules of Tajik.

**Objectives:** To algorithmize the process of morphological analysis Tajik word form (slovoform) and to implement it as a software package

**Research methods:** Set theory, system analysis, elements of coding theory, math modeling, math statistics, object-oriented programming.

**The obtained results and novelty:** The novelty of the dissertation results are:

- the wide base of Tajik morph is formed, which contains 81 prefixes, 76539 roots and 128760 postfixes;
- the classification of types of affixes (inflectional, derivational and word- combinations) is suggested and similar classification of word form (slovoform) also is given;
- the positional coding of Tajik word form (slovoform) is developed;
- an equivalent representation of word form is developed with fragments of the sentence;
- algorithmic support of the morphological analysis of the Tajik word forms is developed.

**Usage degree:** Based on of morphological analyzer language packs for Tajik spellchecking in OpenOfficeOrg and Microsoft Office are created, which are registered as informational resources 3И-03.2.222TJ from 11.01.2012 and № 4201200235 from 04.10.2012. The packs are used in many companies and offices of Tajikistan

**Application:** Solving the tasks of automatic text processing, computer translation, spellchecking, analysis and synthesis of speech, dialogue with computer, automation of indexing processes, annotation, abstracting, classification and rubric of documents, extracting keywords, marking of the hull of the Tajik language.



**Ба чоп имзо шуд 14.05.2018**  
**Формат 6x84/16. Қоғаз офсетӣ**  
**Теооди нашр 100 нусха. Ҳаҷм 1,3 ҷ.ҷ. Фармоиши №161**  
**Дар нашрияи «Меҳвари Дониш» чоп шуд**  
**ш. Хучанд, қўчаи. Ленин, 226**

