

«6D075100 – Информатика, есептеу техникасы және басқару» мамандығы бойынша философия докторы (PhD) ғылыми дәрежесін алу үшін жазылған диссертацияның  
**АНДАТПАСЫ**

**Еркебұлан Гүлнұр Тұратайқызы**  
**«Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесі»**

**Зерттеу тақырыбының өзектілігі.** Тілдер тізімінде ана тілінде сөйлейтіндер саны бойынша ағылшын тілі 3-ші орын (379 миллион адам), орыс тілі – 7-ші орын (154 миллион адам), ал қазақ тілі – 76-шы орын алады (12,9 миллион адам) [1]. Көптілділікті талдауға сәйкес, үш тілде еркін сөйлейтін адамдар саны әлем халқының 13%-ын, ал екі тілде еркін сөйлейтіндер саны 42%-ын құрайды [2]. Қазақстан Республикасында үш тілді білу мансаптық өсу мен жақсы еңбекақы алудың міндетті шарты болып табылады. Қазақстан Республикасының мемлекеттік тілі – қазақ тілі, қазақ тілімен қатар ресми түрде орыс тілі қолданылады. 2014 жылы 1-сыныптан бастап ағылшын тілі оқытыла бастады.

Бірнеше тілді білу орасан зор ақпараттар мен инновациялар ағынымен үлкен жаһандық әлемге терезе ашады. Көптілділік – бұл заман талабы және мемлекет тарапынан бұл даму бағытына үлкен көңіл бөлінуде.

Қазіргі уақытта үш тілде білім беруді дамытудың 2015-2020 жылдарға арналған Жол картасын іске асыру аяқталуда [3], оның қорытындысы бойынша 2023 жылдан бастап ҚР мектептерінде үш тілде білім беруге көшу кестесі әзірленді [4]. 2022-2023 жылдары көптілді мектептерде ағылшын тілінде білім беруге ауысу жоспарланған болса, 2023-2024 жылдары – барлық жалпы білім беретін орта оқу орындарының ауысуы көзделуде. Үштілді білім беруді енгізу білім беру ұйымының педагогикалық кеңесі мен ата-аналар комитетінің алқалық қорытындысы негізінде таңдау бойынша жүзеге асырылатын болады. Үш тілде оқытуға көшу «100 қадам» Ұлт жоспарының 79-қадамын [5] және Қазақстан Республикасының Білім және ғылымды дамытудың 2016-2019 жылдарға арналған мемлекеттік бағдарламасын іске асыру шеңберінде орындалуда [6].

Тіл саласы – бұл үлкен бизнес. TechNavio сарапшылары өздерінің «Global Language Services Market 2020-2024» зерттеуінде 2020-2024 жылдар аралығында CAGR көрсеткіші 4%-ға жететін нарықтың 9,72 миллиард АҚШ долларына өсуін болжайды [7].

Бүгінгі таңда шет тіліндегі бейтаныс сөзді аударуға көмектесетін көптеген онлайн аудармашылар мен браузерлерге арналған кеңейтімдер бар [8]. Бәрінің аудару алгоритмдері әртүрлі, сөздер мен сөз тіркестерінің дерекқорлары әртүрлі, сондықтан аударма нәтижелері де өзгеше болуы мүмкін. Ең жақсы аударманы қалай табуға болады? Ғаламтор желісінде әртүрлі онлайн-компаниялардан бір уақытта бірнеше аударма нұсқаларын ұсынатын MyTranslator атты тек бір ғана браузер кеңейтімі табылды. Қосымша Google Translate, Microsoft Bing Translator, Yandex Translator және DeepL Translator аудармаларын ұсынады, нұсқаларын көру үшін «G», «M», «T» және «Y» қойындыларына ауысу керек. Кеңейтім 2020

жылдың қаңтарында іске қосылған және 10000-нан астам пайдаланушысы бар [9]. Сонымен қатар, мәтінді сөйлемдер бойынша аударғанда әр сөйлем үшін кеңейтімді қайта шақыру керек, аудармашылардың барлық нұсқаларын қойындыларды ауыстырмай көру мүмкіндігі жоқ, аударылған мәтінді кейінірек өздігінен түзету үшін таңдалған аударма нұсқасын қандай да бір жолмен белгілеп қою мүмкіндігі жоқ («Таңдау» - «Көшіру» - «Қою» стандартты әдістерін санамағанда). Диссертациялық зерттеудің «Көптілді мәтіндермен жұмыс жасауда ең сенімді аударманы анықтау» атты бірінші бағытында аталған кемшіліктер жоқ. Сонымен қатар, бұл бағыт пайдаланушыға ең жақсы таңдауды ұсына отырып, әрбір аударма нұсқасын бағалайды.

Мәтінді автоматты түрде құратын интернет ресурстары [10-12] аудармашы сайттарымен бірге [13-15] бір мәтіннің көптеген түсіндірмелерін жасай алады. Түпнұсқаны алу үшін нейрондық желілердің мүмкіндіктерін пайдаланады [16-18].

Бүгінгі таңда нейрондық желілер болжау, жіктеу және басқару мәселелерінде кеңінен қолданылады. Олардың күшті жақтарына «белгісіз заңдылықтары бар мәселелерді шешу, кіріс деректеріндегі шуға төзімділік, нейрондық желіні аппараттық іске асырудағы ықтимал ультра жоғары өнімділік және ақауларға төзімділік» жатады [19]. Сонымен қатар, оның оқуы мен жұмысы үшін үлкен есептеу ресурстары қажет [20].

Диссертациялық зерттеудің екінші бағытында Ренья энтропиясына негізделген мәтіннің аудармасын анықтаудың баламалы әдісі ұсынылған, ол айтарлықтай компьютерлік қуаттылық пен іздеуге көп уақыт қажет етпейді. Ренья энтропиясы даму өзегі ретінде кездейсоқ таңдалмаған. Энтропиялық тәсіл қазірдің өзінде зерттеліп, мәтіндермен жұмыс жасау кезінде қолданылады, бірақ басқа мәнмәтіндерде пайдаланылады. Сонымен, 2019 жылғы [21] жұмысында энтропиялық салмақ әдісін қолдана отырып, құрылымдалған мәтін үшін тұжырымдамаларды алу бойынша зерттеулердің нәтижелері келтірілген, 2020 жылғы [22] жұмысында мәтін ұзындығы мен лексикалық байлық арасындағы энтропиялық байланысты зерттеу нәтижелері келтірілген, 2019 жылғы [23] жұмысында, Войнич қолжазбасының мысалында күмәнді мәтін түпнұсқаларының энтропиялық талдауы зерттелген, 2019 жылғы [24] энтропиялық бағалауды машиналық аудармамен байланыстыратын жұмысында үдерісті екі фазаға бөлу арқылы жеткіліксіз аударманы шешу әдістері ұсынылады. Осы және басқа жұмыстар туралы толығырақ ақпаратты диссертациялық зерттеудің 2.3.2 тармағында табуға болады.

Интернет технологиялары дамып, кеңейіп, қол жетімді және кең тараған сайын, сайттың құндылығы да өседі.

Қазіргі замануи жаһанданған әлемде сайт жаңа келушілерді тарту, оларды өнімдер немесе қызметтер жайлы хабардар ету, сатылым жүргізу қабілетімен бизнес түріне қарамастан, кез келген компания үшін өте құнды актив болып табылады.

«Егер веб-сайттың мазмұны ана тіліңізде, бірақ мазмұнның сапасы төмен болса, сіз сол веб-сайттан сауда жасар ма едіңіз?» деген сұраққа іріктеу тобының

(3000 адам, 50% ерлер / 50% әйелдер 8 елде, шамамен 3 континентте, 25 пен 65 жас аралығындағы және ағылшын тілін білу деңгейі әртүрлі) жауаптарын талдау негізінде 45,3%-ы мазмұны ана тілінде нашар аударылған веб-сайтта сауда жасауға келіспейтіні анықталды [25]. Демек, сайтты локализациялау мәселесі өте өзекті болып келеді, сондықтан университеттерде осы тақырыпқа арналған лингвистикалық және маркетингтік ерекшеліктерін, сонымен қатар әртүрлі техникалық аспектілерін зерттейтін жаңа пәндер пайда болуда [26, 27].

Әрбір оқу орны шетелдік студенттер мен инвесторларды тартуға мүдделі, сондықтан университеттің веб-сайты мемлекеттік және жеке ұйымдардың веб-сайттары сияқты халықаралық ынтымақтастық үшін тиісті өкілдік болуы керек. Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесінде жетіспейтін ақпараттың машиналық аударма жасайтын мүмкіндігі бар, көптілді сайттың барлық тілдік нұсқаларындағы жетіспейтін аудармаларын анықтайтын тағы бір бағыт қарастырылады.

Жаңалықтар агрегаторлары да көптілді сайттар болып табылады, бірақ диссертациялық зерттеуде олар үшінші емес, төртінші бағытта қарастырылады. Бұл алгоритмнің жарияланған материалдарға емес, талдаудан кейінгі және жариялау алдындағы кезеңдегі ақпаратты құруға қолданылатындығына байланысты.

Бірнеше тілге арналған тесттер мен оқу құралдарын жасауға арналған соңғы бағыт, әртүрлі тілдердегі ақпараттың шынайылығының маңыздылығын қарастырады. Ақпараттың шынайылығы – аударма мен түпнұсқаның бірдей екендігіне кепілдік беретін қасиет. Егер материалдар түпнұсқаға сәйкес болса, аударма тіліндегі материалдар жақсы игерілетінін көрсеткен ғылыми тәжірибелер бар [28,29]. Аударманың дұрыстығын автоматты режимде тексеретін бағдарламаны табу қиын. Осы уақытқа дейін онлайн-аудармашыларда қателер немесе кейбір сөздер мен сөз тіркестерінің нұсқаулары қазақ тілінде жоқ сондықтан аудармаларды тексеру үшін кейде сарапшылар тобы қосарланған тексеруді қолданады.

Тест тапсырмаларын құрастыру бойынша әдістемелік ұсынымдарда [30, 31] алтыншы тараудың 6.5 – тармағында: «Мемлекеттік, орыс және басқа тілдерде тест тапсырмаларын әзірлеу, жаңарту кезінде – тест тапсырмаларын аударғанда сұрақтар мен жауаптар шынайы болып, ҚР Үкіметі жанындағы Мемлекеттік терминологиялық комиссиясы бекіткен терминологиялық сөздіктерге сәйкес болуы тиіс» деп жазылған. Диссертациялық жұмыста синонимдер мен терминологиялық сөздіктер сериялары негізінде тесттер мен оқу құралдарының аудармаларын, сондай-ақ түпнұсқа мәтіннің жетіспейтін немесе қате аударылған бөліктерін анықтау үшін жалпы ақпаратты талдауды қалыптастыру кезінде автоматты түрде тексеру мен кеңес беру мүмкіндігін қарастыру ұсынылады.

Нәтижесінде «Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесі» (бұдан әрі – КМҮСЖ) диссертациясында зерттеудің бес негізгі бағыты анықталды:

1) онлайн-аудармашылар арасында көптілді мәтіндермен жұмыс істеудегі ең дұрыс аударманы анықтау;

2) энтропиялық тәсіл негізінде түпнұсқа мәтіннің шынайы және жалған аудармасының айырмашылығын анықтау;

3) көптілді сайттың барлық тілдік нұсқаларында жетіспейтін аудармаларын анықтау;

4) көптілді жаңалықтар агрегаторларының жұмысындағы жетіспейтін аудармаларды анықтау;

5) бірнеше тілге арналған тест және оқу-әдістемелік құралдарын жасау.

Жоғарыда аталған фактілердің барлығы көптілді мәтіндермен жұмыс жасауда аударма мен түпнұсқа мәтіндерін сәйкестендіру өзекті екенін көрсетеді.

**Зерттеудің мақсаты:** «100 қадам» Ұлт жоспарының 79-қадамын іске асыру шеңберінде «сөйлем» және «параграф» үлгілерін пайдалана отырып, көптілді мәтіндерге байланысты бағыттарда, Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесін басқару моделін және алгоритмдерін әзірлеу.

Мақсатқа жету үшін келесі зерттеу тапсырмалары қойылды:

1. КМҮСЖ-де пайдалану үшін плагиатты анықтау жүйелері мен онлайн аударма қызметтерін талдау, КМҮСЖ -нің негізгі бағыттарын анықтау үшін деректерді жүйелеу.

2. Бағдарламалық жүзеге асыру және сарапшылық бағалау негізінде мәтінді және оның аудармасын (онлайн аудармашылардан) салыстыру үшін қатарларды анық емес салыстырудың нақты алгоритмін таңдау үшін есептеулерді жүргізу. Есептеулер үшін орыс, қазақ, ағылшын тілдеріндегі параллель мәтіндер корпусын құру.

3. Таңдалған қатарларды анық емес салыстыру алгоритмін қолдана отырып, онлайн аудармашылар арасында көптілді мәтіндермен жұмыс істеу кезінде ең сенімді аударманы анықтау және бірнеше тілдерге арналған тесттер мен оқу құралдарын құру кезінде қателердің алдын алу алгоритмін жасау.

4. Координаттарды есептеуді бағдарламалық қамтамасыз ету арқылы, әртүрлі тілдердегі мәтіндердің жақындығын анықтауға арналған энтропиялық тәсілдің жұмыс қабілеттілігін тексеру үшін есептеулер орындау (мәтіндердің жоғары жиіліктегі сөздерінің негізгі қатарларын қалыптастыру, «сөйлем» және «параграф» үлгілері үшін энтропиялық координаттарды есептеу, Минковский метрикасына сәйкес мәтіндер энтропиясының жиынтығы арасындағы қашықтықты есептеу).

5. Түпнұсқа мәтіннің шынайы және жалған аудармасының айырмашылығын анықтау алгоритмін, көптілді сайттың барлық тілдік нұсқаларында жетіспейтін аудармаларын анықтау алгоритмін және энтропиялық тәсілді қолдана отырып, көптілді жаңалықтар біріктіруші жұмысындағы жетіспейтін аудармаларды анықтау алгоритмін әзірлеу.

6. Жоғарыда аталған зерттеулердің нәтижелеріне жүктеле отырып, Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесінің басқару моделін жасау.

**Зерттеу нысаны** әртүрлі тілдердегі мәтіннің маңызды ақпараттық блоктарын қамтитын ақпараттық өріс болып табылады.

**Зерттеу пәні** – көптілді мәтіндердің үлгілерін сәйкестендірудің басқару моделі мен алгоритмдері болып табылады.

**Зерттеу әдістері:** Оливер қатарын анық емес салыстыру алгоритмі, FuzzyWuzzy қатарын анық емес салыстыру алгоритмі, Портер стеммерлері, қалыпқа келтіру, Шеннон энтропиясы, Реньи энтропиясы, Минковский метрикасы, Хэмминг қашықтығы, декарттық қашықтық, масса орталықтары арасындағы қашықтық, геометриялық орталықтар арасындағы қашықтық, параметрлік орташа орталықтар арасындағы қашықтық.

**Ғылыми жаңалық:** Ғылыми жаңалық белгілі php-шешімдері мен параметрленетін энтропия, және көптілділік ерекшеліктерін ескере отырып, мәтіндік материалдардың үлгілерін сәйкестендіру әдістерін түзу нәтижесі ретінде автордың ұсынған КМҮСЖ дәлелдеулерінен тұрады.

**Алынған нәтижелердің жаңалықтары:**

- орыс, қазақ, ағылшын тілдерінде параллель мәтіндер корпусы әзірленді;
- онлайн аудармашылар арасында көптілді мәтіндермен жұмыс істеу кезінде ең сенімді аударманы анықтау алгоритмі және анық емес қатарды салыстыру алгоритмін қолдана отырып, бірнеше аударма тілдеріне арналған тесттер мен оқу құралдарын жасау алгоритмі әзірленді;

- көптілді мәтіндердің жақындығын анықтау үшін энтропикалық әдіс әзірленді;

- түпнұсқа мәтіннің шынайы және жалған аудармалары арасындағы айырмашылығын анықтау алгоритмі, көптілді сайттың барлық тілдегі нұсқаларында жетіспейтін аудармаларын анықтау алгоритмі және энтропиялық тәсілді қолдана отырып көптілді жаңалықтар агрегаторлары жұмысындағы жетіспейтін аудармаларды анықтау алгоритмі жасалды;

- «100 қадам» Ұлт Жоспарының 79-қадамын жүзеге асыру аясында КМҮСЖ-ін басқару моделі әзірленді.

**Теориялық маңыздылығы:** негізгі теориялық жаңалықтар авторлық энтропиялық тәсілдердің әртүрлі тілдердегі мәтіндердің жақындығын анықтау мүмкіндігін табумен қатар көптілді мәтіндермен жұмыс істеу кезінде қатарларды анық емес салыстыру алгоритмін тәжірибе жүзінде анықтау болды. Мәтіндік корпусы құру бойынша ұсыныстар онлайн аудармашыларды қолданатын, көптілді мәтіндер саласын зерттеушілерге пайдалы болмақ. Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесінің басқару моделі мен алгоритмдері, халықтың көптілділігіне байланысты туындаған мәселелердің бірқилы шешімдерінің сипаттамасы болып табылады.

**Тәжірибелік маңыздылығы:** Жұмыстың практикалық маңыздылығы әзірленген жүйенің көптілділікті ескере отырып, ақпараттық өріспен өзара әрекеттесетін ұйымдардың, сонымен қатар кез келген қызығушылық танытқан пайдаланушы үшін де қолайлы болып табылады. Себебі бұл жүйенің бағдарламалық элементтері ғаламторда қоғамға қолжетімді жерде орналастырылған. Берілген үлгілерді пайдалану мүмкіндігі бар алгоритмдерді қажетті скриптерді ғаламторда қоғамдық доменде орналастыру арқылы (бірнеше онлайн-аудармашылардың аударма нұсқалары негізінде, әдетте, бағдарламаны пайдаланғаны үшін төлемді талап етеді), аудармашылардың, талдаушылардың

және басқа да мүдделі пайдаланушылардың көптілді мәтіндермен жұмыс істеу барысында ең сенімді аударманы анықтауы үшін қолдану мүмкіндігі бар.

Бастапқы мәтінге сәйкес келетін аударманы анықтауға қатысты параметрленген энтропиясы бар әзірленген алгоритмдерді ұсынуға болады:

- аналитикалық компанияларға (мақаланың/жаңалықтардың түпкі деректемелерін іздеу, мақаланы/жаңалықтарды авторлық жұмыс пен кірме бөліктеріне бөлу және т.б.);

- ақпараттық қауіпсіздік саласындағы ұйымдарға (телнұсқаларды іздеу, ұлттық қауіпсіздікке қатер төндіретін басқа тілдегі материалдардың түпкі деректерін іздеу және т.б.);

- жоғары оқу орындарына (студенттердің жұмыстарынан аударылған плагиатты іздеу) және т.б.

Көптілді жаңалықтар агрегаторлары жұмысындағы жетіспейтін аудармаларды анықтау және көптілді сайттың барлық тілдік нұсқаларында аудармаларын анықтау бағыттарында әзірленген алгоритмдерді ең алдымен ақпарат агенттіктерінің иелері үшін, сондай-ақ мемлекеттік және жекеменшік көптілді интернет ресурстарында қолданылуға болады.

Ұсынылған алгоритмді әртүрлі тілдерде тесттер мен оқу құралдарын жасау кезінде көлемді мәтіндік материалдарды өңдеумен айналысатын көптеген мүдделі адамдар қолдана алады.

#### **Қорғауға ұсынылған диссертацияның ережелері (ғылыми нәтижелер):**

- онлайн аудармашылар қызметтерін пайдалана отырып, көптілді мәтіндердің үлгілерін іздеу жолдарын зерттеу нәтижесінде мәтіндік корпус құру бойынша ұсыныстар;

- көптілді мәтіндердің жақындығын анықтаудың энтропикалық тәсілі;

- көптілді мәтіндер үшін үлгіні сәйкестендіру жүйесінің басқару моделі мен алгоритмдері;

- мәтіннің үлгілерін салыстыру және оның аудармасын қатарларды анық емес салыстыру арқылы есептеудің бағдарламалық іске асыру, сонымен қатар энтропия координаттарын есептеуді бағдарламалық қамтамасыз ету.

**Автордың жеке үлесі** қорғауға шығарылатын негізгі ережелерді дәлелдейтін зерттеулер жүргізу мен алынған нәтижелерді талдау және жинақтау кезіндегі рөлдің маңыздылығынан тұрады.

**Диссертацияның құрылымы мен көлемі.** Диссертацияның классикалық құрылымы : кіріспе бөлім, негізгі бөлім (үш тарау), қорытынды, пайдаланылған деректер тізімі және қосымшалардан тұрады. Жұмысқа 65 сурет, 14 кесте және пайдаланылған дереккөздердің 113 атауы кіреді.

**Кіріспеде** зерттеу тақырыбын таңдау негізделген, КМҮСЖ-нің бес бағытының өзектілігі ашылған, оның міндеттерін нықтай отырып, зерттеу мақсаты тұжырымдалған, зерттеу объектісі мен пәні ұсынылған, жұмыстың ғылыми жаңалығы, практикалық маңыздылығы ашылған.

**Бірінші тарауда** Рунеттегі плагиатты анықтау жүйелері салыстырылған. «Антиплагиат. Вуз» жүйесінің аударылатын кірме сөздерін іздеу модулінің жұмысына толық талдау жасалды. Google және Yandex сервистерінің көмегімен

көптілді мәтіндердің үлгілерін іздеу әдістері ашылған. Диссертациялық зерттеудің міндеттерінің тұжырымы жасалды.

**Екінші тарауда** КМҮСЖ үлгілері мен әдістері ұсынылған. Php коды арқылы ең сенімді аударманы анықтау бағыты зерттелді. Бес кезеңнен тұратын мәтіндердің жіктелуі жан-жақты сипатталған. Мәтіндердің үйлесімділік дәрежесін бағалау ретінде энтропиялық тәсілді қолдану қарастырылды. Диссертациялық зерттеу есептері үшін энтропиялық тәсілдің тиімділігі мен КМҮСЖ-нің бағыттары үшін қатарды анық емес салыстыру алгоритмін таңдау тәжірибе жүзінде дәлелденді. Тәжірибиелер нәтижесінде скриптер жасалды, оларға авторлық құқықпен қорғалатын объектілерге құқықтардың мемлекеттік тізіліміне мәліметтерді енгізу жайлы екі куәлік алынды (Диссертацияның D, E қосымшалары).

Жүргізілген тәжірибиелер барысында анықталған қатарды анық емес салыстырудың ең жақсы алгоритмі КМҮСЖ-нің бірінші және бесінші бағыттарында, ал жасалған энтропиялық тәсіл – екінші, үшінші және төртінші бағыттарында қолданылды.

**Үшінші тарау** екінші тарауда келтірілген тәсілдер мен әдістемелерді қолдана отырып, КМҮСЖ-нің бес бағытының басқару моделі мен алгоритмдеріне арналған. Үшінші тарауда мәтіндік үлгілерді салыстыру және қатарларды анық емес салыстыру арқылы аудару есептеулерінің орындауын бағдарламалық қамтамасыз ету, сонымен қатар энтропия координаттарын есептеудің бағдарламалық қамтамасыз ету (авторлық құқықтар алынған скриптер) толық сипатталған.

**Қорытындыда** диссертациялық зерттеу нәтижелеріне сүйенетін негізгі қорытындыларды қамтитын зерттеу нәтижелері ұсынылған.

**Жұмысты апробациялау.** Диссертациялық зерттеудің нәтижелері ғылыми конференцияларда баяндалып, талқыланды:

- «GLOBAL SCIENCE AND INNOVATIONS 2019: CENTRAL ASIA» VII Халықаралық ғылыми-практикалық конференция Нұр-Сұлтан қаласы;
- Варшавадағы халықаралық конференция MODERN SCIENTIFIC CHALLENGES AND TRENDS (2019 г.) басылым аясында;
- «Заманауи университет цифрлық ойлау кеңістігі ретінде» халықаралық ғылыми-әдістемелік конференциясы Новосибирск қ. (2020 г.).

Ғылыми тағылымдама жүргізілді. Авторлық құқықпен қорғалатын объектілерге құқықтардың мемлекеттік тізіліміне мәліметтерді енгізу туралы 2 анықтама алынды.

**Басылымдар мен авторлық куәліктер.** Негізгі зерттеу нәтижелері 8 ғылыми мақалада жарияланған, оның ішінде халықаралық рецензияланған ғылыми журналдарда (Scopus) 2 мақала жарияланған, ғылыми қызметтің негізгі нәтижелерін жариялауға ұсынылған ғылыми жарияланымдар тізіміне енгізілген ғылыми журналдарда 3 мақала, уәкілетті өкілді ұйымымен бекітілген, 3 жұмысы – халықаралық ғылыми конференциялар жинағында. 2 авторлық куәлік алынған.

### **Ғылыми жарияланымдар тізімі:**

1. Еркебұлан Г.Т., Куликова В.П. «Антиплагиат» жүйесінде аударылған кірме сөздерді іздеуді жүзеге асырудың ерекшеліктері: күшті және әлсіз жақтары//Матер. халықарал. ғылыми. конф. «Modern scientific challenges and trends» Polish Science Journal. – Варшава, 2019. – № 9(20). – Б. 42-46.

2. Еркебұлан Г.Т., Куликова В.П. Іздеу жүйелері көптілді мәтіндерде кірме сөздерді анықтау ретінде // Global Science and Innovations 2019: Central Asia. – Нұр-Сұлтан, 2019. – № 2(3). – Б. 171-173.

3. Еркебұлан Г.Т., Куликова В.П. Кросс-тілдік (аударма) плагиатты анықтау жүйесін салыстырмалы талдау // Қ. Сатпаев атындағы ҚазҰТЗУ Хабаршысы – Алматы, 2019. – № 6(136). – Б. 178-183.

4. Еркебұлан Г.Т., Куликова В.П., Куликов В.П. Кросс-тілдік плагиатты анықтауда Google Custom Search және Google Translate API қолдану туралы // АЭЖБУ Хабаршысы. «Ақпараттық технологиялар» сериясы». – Алматы, 2019. – № 4(47). – Б. 109-116.

5. Еркебұлан Г.Т., Куликова В.П., Куликов В.П. Яндекс қолдану туралы. XML және API Яндекс. Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесіндегі аудармашы// АЭЖБУ Хабаршысы. «Ақпараттық технологиялар» сериясы». – Алматы, 2020. – № 1(48).– Б. 110-117.

6. Еркебұлан Г.Т., Куликова В.П., Куликов В.П., Крылова Е.М. Көптілді мәтіндердің үлгілерін сәйкестендіру жүйесіндегі мәтіндік сұраныстарды жіктеудің модельдері мен әдістері // Матер. халықарал. ғылыми. конф. «Қазіргі университет сандық ойлау кеңістігі ретінде». - Новосибирск: СГУГиТ, 2020. – Б. 130-134.

7. G. Yerkebulan, V. Kulikova, V. Kulikov. Google/Yandex Translation Detection in the Patterns Identifying System of Multilingual Texts // Research Institute for Intelligent Computer Systems, West Ukrainian National University, журнал «International Journal of Computing», ISSN 1727-6209 (print). – 2021. – Vol. 20, Issue 1. – P. 72-77. // <https://doi.org/10.47839/ijc.20.1.2094>

8. G. Yerkebulan, V. Kulikova, V. Kulikov, Z. Kulsharipova. Devising an entropy based approach for identifying patterns in multilingual texts // PC TECHNOLOGY CENTER, Kharkiv, Ukraine, журнал «Eastern-European Journal of Enterprise Technologies», ISSN 1729-3774 (print). – 2021. – №2/2 (110). – P. 16-22.// <https://doi.org/10.15587/1729-4061.2021.228695>

### **Авторлық құқықпен қорғалатын объектілерге құқықтардың мемлекеттік тізіліміне мәліметтерді енгізу туралы куәліктер:**

1. 2021 жылдың «17» тамызынан №19775 авторлық құқықпен қорғалатын объектілерге құқықтардың мемлекеттік тізіліміне мәліметтерді енгізу туралы куәлік «Көптілді мәтіндермен жұмыстағы ең сенімді аударманы анықтауға арналған бағдарлама» (ЭЕМ-ге арналған бағдарлама) авторлары: Еркебұлан Г. Т., Куликова В. П., Куликов В. П.

2. 2021 жылдың «30» шілдесінен №19562 авторлық құқықпен қорғалатын объектілерге құқықтардың мемлекеттік тізіліміне мәліметтерді енгізу туралы куәлік «Көптілді мәтіндердің жақындығын анықтау үшін энтропиялық



координаттарды есептеу бағдарламасы («параграф» және «ұсыныс» үлгілері бойынша)» (ЭЕМ-ге арналған бағдарлама) авторлары: Еркебұлан г. Т., Куликова В. П., Куликов В. П.