«6D060200 – Информатика» мамандығы бойынша философия докторы (PhD) дәрежесін алу үшін ұсынылған ізденуші Тойганбаева Назгуль Абеновнаның «Қазақ-орыс тілінің қолжазба мәтінін оптикалық тану үшін құралдар әзірлеу және деректер базасын құру» тақырыбына орындаған диссертациялық жұмысына ғылыми жетекшісі ф-м.ғ.к., қауымдастырылған профессор Нурсеитов Данияр Борисовичтің

# ПІКІРІ

Ізденуші Тойганбаева Назгуль Абеновнаның «Қазақ-орыс тілінің қолжазба мәтінін оптикалық тану үшін құралдар әзірлеу және деректер базасын құру» тақырыбындағы ғылыми жұмысы қазіргі ғылым мен технологияның даму үрдісіне сай келетін, қолданбалы мәні зор өзекті мәселені қамтиды. Жұмыс қазақ және орыс тіліндегі қолжазба мәтіндерді оптикалық тану мәселесін шешуге бағытталған.

Зерттеу аясында алғаш рет қазақ тіліндегі сөз деңгейіндегі қолжазба мәтіндер корпусы Kazakh Offline Handwritten Text Dataset (KOHTD) жасалып, ол осы саладағы маңызды ресурстардың біріне айналып отыр. Сонымен бірге, қазақ-орыс тіліндегі қолжазбаларды оптикалық тану үшін нейрондық модельдер қазақ тіліне бейімделіп оқытылып, нақты нәтижелер алынған.

Аталған нәтижелер қазақ және орыс тіліндегі қолжазбаларды тануға арналған әдістер мен құралдардың жоқтығын ескере отырып, бұл зерттеудің ғылыми жаңалығын және өзектілігін айқындай түседі. Жұмыстың нәтижелері мемлекеттік архивтерді, білім беру контентін және тарихи құжаттарды цифрландыру, сондай-ақ қазақ тілінің цифрлық кеңістіктегі қолданысын кеңейту ісіне елеулі үлес қосады.

Зерттеу барысында келесі нәтижелер алынды:

- Қолжазба мәтіндердің деректерін жинау, алдын ала өңдеу, анотациялау және корпусқа енгізу үдерістерін автоматты түрде орындайтын толығымен автоматтандырылған жүйе жасалды және тәжірибелік негізде іске қосылды.
- Қазақ тіліндегі қолжазба мәтіндерден сөздерді дәл сегменттеуге арналған генетикалық алгоритм әзірленіп, оның эвристикалық мүмкіндіктері арқылы әріпаралық шекараларды анықтау дәлдігі арттырылды.
- Қазақ тіліндегі оффлайн қолжазба мәтіндерден тұратын Kazakh Offline Handwritten Text Dataset (KOHTD) атты ашық датасет жасалды.
- Қазақ-орыс тілдеріндегі қолжазба мәтіндерді оптикалық тануға арналған Flor, Puigcerver, Bluche және Scheidl нейрондық модельдері қазақ тілінің графикалық ерекшеліктеріне байланысты икемделіп оқытылды.

Зерттеу тақырыбы бойынша алынған нәтижелер ғылыми басылымдарда, оның ішінде Scopus базасына енген журналда 1 мақала, ҚР ҒЖБ саласындағы сапаны қамтамасыз ету комитеті ұсынған ғылыми басылымдар тізбесіндегі журналдарда 3 мақала, сондай-ақ халықаралық конференция материалдарында 1 мақала жарияланды. Зерттеу нәтижелері әл-Фараби атындағы ҚазҰУ

Сканировано с CamScanner

компьютерлік ғылымдар кафедрасында ғылыми семинарларда, сонымен қатар, халықаралық конференцияларда баяндалды.

Диссертациялық жұмыста қойылған мақсаттар мен міндеттер толық көлемде орындалған. Автор заманауи әдістер мен алгоритмдерді тиімді қолданып, ғылыми жаңалығы бар әрі практикалық маңызы жоғары нәтижелерге қол жеткізген. Жасалған зерттеу қазақ және орыс тіліндегі қолжазба мәтіндерді тану саласында отандық ғылым үшін маңызды үлес болып табылады. Әзірленген технологиялар мен ресурстар болашақта ақпараттық жүйелерді дамытуда, цифрландыру үдерісін жеделдетуде, сондай-ақ қазақ тілінің сандық қолданыс аясын кеңейтуде қолданылуы мүмкін. Жұмыс құрылымы мен мазмұны, ғылыми тілдің сақталуы және зерттеу логикасы жоғары деңгейде орындалған.

Жалпы алғанда, диссертациялық жұмыс ғылыми жаңалығы, теориялық тереңдігі және практикалық құндылығымен ерекшеленеді. Тойганбаева Назгуль Абеновнаның «Қазақ-орыс тілінің қолжазба мәтінін оптикалық тану үшін құралдар әзірлеу және деректер базасын құру» тақырыбындағы диссертациялық жұмысы 6D060200 – Информатика мамандығы бойынша философия докторы (PhD) ғылыми дәрежесін алу талаптарына толық сәйкес келеді және қорғауға ұсынамын.

**Ғылыми жетекші,**
**«ҚМГ Инжиниринг» ЖШС сарапшысы,**
**ф-м.ғ.к.,**
**қауымдастырылған профессор** «ҚМГ Инжиниринг» ЖШС **Нурсеитов Д.Б.**

**«ҚМГ Инжиниринг» ЖШС**
**адами ресурстарды басқару**
**департаментінің директоры** **Шуманах Г.Е.**

Сканировано с CamScanner