

Касекеева Айслу Бисеновнаның
«8D061- Ақпараттық-коммуникациялық технологиялар»
кадрларды даярлау бағыты
«6D070300 – Ақпараттық жүйелер» мамандығы бойынша философия
докторы (PhD) дәрежесін алуға ұсынылған «Ағылшын, орыс және қазақ
тілдерінде ақпараты бар семантикалық кітапхананың моделі»
диссертациялық жұмысының

АНДАТПАСЫ

Цифрлық сауаттылық-ақпараттық қоғамдағы қауіпсіздіктің негізі, ХХІ ғасырдың аса маңызды білімінің, бүгінгі қоғамның негізгі тақырыптарының бірі. Цифрлық сауаттылық – бұл адам өмірінің барлық салаларында цифрлық технологияларды сенімді және тиімді пайдалануға әзірлік пен қабілеттілік. Бұл технологияны қолдану халықтың өмір сүру сапасын жақсартуға мүмкіндік береді.

2017 жылғы 31 қаңтардағы "Қазақстанның Үшінші жаңғыруы: жаһандық бәсекеге қабілеттілік" Жолдауында Президент Нұрсұлтан Назарбаев "Цифрлық Қазақстан" мемлекеттік бағдарламасын әзірлеу қажеттігін атап өтті: "... Біз цифрлық технологияларды пайдалана отырып құрылған жаңа салаларды дамытуға тиіспіз. Бұл – маңызды кешенді міндет. Елімізге 3D-баспа, интернет-дүкендер, мобильді банкинг, цифрлық қызметтер, денсаулық сақтау, білім беру және басқа да перспективалы бағыттар сияқты бағыттарды дамыту қажет. Бұл салалар қазірдің өзінде дамыған елдердің экономикасын қайта құрылымдауды жүргізіп, дәстүрлі салаларға жаңа сапа берді. Осыған байланысты Үкіметке "Цифрлық Қазақстан" жеке бағдарламасын әзірлеуді және қабылдауды тапсырамын. ... Цифрлық индустрияның дамуы барлық басқа бағыттарға серпін береді. Сондықтан Үкімет ІТ индустриясын дамытуға ерекше назар аударуы керек".

Әр ел экономикалық дамуға және әлемдік кеңістікке интеграциялануға әсер ететін белгілі бір факторларды таңдайды. Қазақстан Республикасы әлемдік экономиканың талаптарына жауап беретін, халықтың өмір сүру деңгейі мен жалпы әл-ауқатын арттыратын білім берудің негізгі факторларының бірі болып табылады. Көптеген елдердің үкіметтері білім беру қызметтерінің сапасын дамыту арқылы экономиканың бәсекеге қабілеттілігін арттыруды басты міндеттердің бірі ретінде қарастырады. Бұл қоғамның білім деңгейі мен ғылыми әлеуеті экономикалық өсудің маңызды шарты болып табылатындығына байланысты. Қазіргі жаһандану процестері аясында ұлттық өзара байланысты жүйелердің жиынтығы болып табылатын халықаралық жоғары білім беру жүйесі қалыптасуда.

Білім - «Қазақстан - 2030» Стратегиясының маңызды критерийлерінің бірі. Білім беру бағдарламаларының басты мақсаты - білім беру жүйесін жаңа экономикалық ортаға бейімдеу. Қазақстан Республикасының Президенті республиканы әлемнің бәсекеге қабілетті 50 елінің қатарына қосу міндетін қойды. Сондай-ақ, «Қазақстан Республикасының ұлттық қауіпсіздігі туралы»

Қазақстан Республикасының Заңына сәйкес, қауіптің бірі - білім сапасы мен елдің интеллектуалды әлеуетінің нашарлауы, бұл құрудың орасан зор маңыздылығын дәлелдейді және сапалы білім беру жүйесін жетілдіреді.

Соңғы уақытқа дейін цифрлы кітапханаларды қарапайым қолданушылар дәстүрлі кітапханалық каталогтардың физикалық кітапхана объектілерінің (әдетте кітаптар немесе басқа баспа басылымдары) сипаттамаларын қамтитын электронды нұсқалары ретінде қабылдады. Заттардың мазмұнын, мазмұнын және құрылымын анықтау осындай кітапханалардың қосымша, бірақ факультативті функциялары ретінде қарастырылады және қабылданады. Интернет пен семантикалық технологиялардың дамуы өзіндік түзетулер енгізеді және сандық кітапханалар ұғымына кеңірек көз жүгіртуге және кітапханалардың жаңа типін қалыптастыру үшін білімнің әр түрлі салаларында ақпараттық жүйелерді енгізудегі жинақталған тәжірибені қорытуға мүмкіндік береді.

Интернеттің қарқынды дамуы жағдайындағы кітапхана ұғымының өзі мүлдем басқа мағынаға ие және пайдаланушының кітапханалар ұсынатын процестерге белсенді қатысуын білдіреді. Мұндай кітапхана пайдаланушылардың осы нақты пайдаланушыға қажетті кітапхана мазмұнын құру, іздеу және жіктеу процесіне қатысуын болжайды.

Зерттеу тақырыбының өзектілігі. Мәтіндік ақпарат көлемінің тез өсуіне байланысты табиғи тілдегі есептеу лингвистикасы саласындағы зерттеулер өзекті болып қала береді. Қазіргі кезде адамдар мен машиналар табиғи тілде көбейтетін ақпарат мөлшері айтарлықтай өсті. Алгоритмдер құру және мәліметтер жинау жүйелерін құру, әртүрлі тілдердегі ақпараттарды қамтитын семантикалық кітапханалар моделі, кітапханаларды мәтіндік құжаттармен жіктеу және кластерлеу әлі де күрделі міндеттер болып табылады.

Көптеген зерттеушілер олардың мағыналық бейнелерін жасау үшін мәтіндерге терең семантикалық талдау жүргізу қажеттілігіне бейім, олардың негізінде құжаттардың жақсы рейтингісін жүргізуге болатын еді. Бұл тәсіл сөзсіз ең ақылға қонымды, бірақ мәтінді автоматты түрде өңдеуге қолайлы құралдарды жасау үшін мұқият әрі ұзақ жұмыс қажет. Атап айтқанда, әртүрлі сараптама салаларының егжей-тегжейлі сипаттамасы қажет болуы мүмкін. Сондықтан ішінара шешімдер іздеудің де мағынасы бар, оның бірі осы жұмыста көрсетілген.

Мәтіндік ақпарат ағынының қарқындылығының үздіксіз артуы әр түрлі тілдердегі ақпараттарды қамтитын кітапхананың семантикалық моделінің міндетін маңызды ете түседі.

Семантика - тіл бірліктерінің мағынасын зерттейтін лингвистиканың бөлімі: жеке сөздер, сөз тіркестері, сөйлемдер, мәтін фрагменттері. Қазіргі уақытта оператор мәндерін ұсынудың бірнеше машинаға бағытталған әдістері бар.

Мысалы, И.А. Мельчук лексикалық функция ұғымын енгізіп, синтаксистік және семантикалық валенттілік ұғымын дамытып, оны түсіндірме-комбинаторлық сөздік аясында қарастырды. В.Ш. Рубашкин және

Д.Г. Лахути семантикалық анализатордың тиімді жұмыс істеуі үшін синтаксистік байланыстар иерархиясын енгізді.

Мәтіндік ақпарат ағынының қарқындылығының үздіксіз артуы әр түрлі тілдердегі ақпараттарды қамтитын кітапхананың семантикалық моделінің міндетін маңызды ете түседі.

Теориялық негізі сілтемелердің грамматикасы, табиғи тілдегі мәтіндердің синтаксистік анализаторлары, сөйлемдерді салыстыру және мәтіндер тақырыбын анықтау әдістері, графиктер мен алгоритмдер бойынша математикалық логикаға арналған зерттеулерден тұратын ғылыми жұмыстар болды.

Диссертациялық зерттеудің мақсаты және ғылыми нәтижелері

Диссертациялық зерттеудің мақсаты - жаратылыстану тіліндегі уақыттық және кеңістіктік құрылымдарды талдау барысында алынған нәтижелер мен тұжырымдарға сүйене отырып, ағылшын, орыс және қазақ тілдеріндегі ақпараттарды қамтитын семантикалық Semantics кітапханасының әдісі мен моделін жасау.

Зерттеу нысаны - ғылым және білім беру қызметін қолдаудың мағыналық кітапханасы.

Зерттеу пәні - модельдер, алгоритмдер және семантикалық кітапханалардың әдістері, сонымен қатар сандық кітапханалардағы ресурстарды жіктеу және кластерлеу әдістері.

Ғылыми жаңалығы. Ең көп ғылыми қызығушылық тудыратын сұрақтар: ағылшын, орыс және қазақ тілдеріндегі сөйлемдердегі кеңістік-уақыттық құрылымдардың жіктелуі ұсынылды. Математикалық модель ұсынылған эквиваленттік тілдегі көптеген ресми көріністер.

Қазақ, орыс және ағылшын. Кеңістік пен уақытқа қатысты жаратылыстану тілінің классификациясы және қазақ, орыс және ағылшын тілдерінің жай сөйлемдерді қамтитын семантикалық эквиваленттілік әдісі мен алгоритмі ұсынылған. Қазақ, орыс және ағылшын тілдерінің семантикалық балама сөздер мен сөз тіркестерін қолдана отырып, семантикалық кітапхананың моделі мен прототипі жасалды.

Link Grammar Parser және Диалинг бағдарламалық жүйелері арқылы уақытша кеңістіктік сөйлемдердің эквиваленттік алгоритмі жүзеге асырылды. Ресми көріністердің моделін ескере отырып, ағылшын, орыс және қазақ тілдеріндегі ақпаратты қамтитын семантикалық кітапхананың ақпараттық жүйесі құрылды.

Link Grammar Parser және Диалинг анализаторлары, PHP серверлік бөлігі, HTML, CSS, MySQL дерекқорының клиенттік бөлігі арқылы кеңістіктік-уақыттық қатынастардың эквиваленттік алгоритмі жүзеге асырылды. Формальді жіктелу моделін ескере отырып, ағылшын, орыс және қазақ тілдеріндегі ақпаратты қамтитын семантикалық кітапхананың ақпараттық жүйесі құрылды.

Зерттеудің міндеттері. Алға қойылған мақсатқа сәйкес диссертациялық жұмыста келесі міндеттер шешіледі:

1. Кеңістік-уақыт конструкцияларын талдау және жіктеуді жүргізу;

2. Эквиваленттік тілде сөйлемдердің формальды көріністерінің жиынтығының математикалық моделін жасау;

3. Semantics семантикалық кітапханасының ақпараттық жүйесінің прототипін жасау.

Қорғауға ұсынылатын мәселер:

1. Кеңістік-уақыт конструкцияларының формальды жіктелуі.

2. Эквиваленттік тілдегі формальды өкілдіктер жиынтығының математикалық моделі.

3. Semantics семантикалық кітапханасының ақпараттық жүйесінің бағдарламалық прототипі.

Алынған нәтижелердің тәжірибелік маңыздылығы. Алынған нәтижелер мен бағдарламалық қосымша интеллектуалды ақпараттық іздеу жүйелерінде қолданыла алады. Атап айтқанда, мәтіннің іздеу сұранысына сәйкестігін анықтау. Бұл лингвист ғалымдарды да қызықтырады.

Ұсынылған диссертациялық жұмыстың негізгі ережелерінің практикалық маңыздылығы Ресей Ғылым академиясының Сібір бөлімшесі А. П. Ершов атындағы информатика жүйелері институтында семантикалық кітапхананы дамыту үшін әзірленген модельдер, әдістер мен алгоритмдерді қолдану нәтижелерімен расталады.

Автордың жеке үлесі. Ағылшын, орыс және қазақ тілдеріндегі ақпараттан тұратын семантикалық кітапхананың модельдері мен әдістерін автордың өзі ұсынды, сипаттады және әзірледі. С.И. Ожеговтың түсіндірме сөздігінен және басқа дереккөздерден кеңістіктік-уақыттық қатынастарға қатысты маңызды ұғымдарды талдау және іріктеу, семантикалық кітапханада үш тілде қолдануға арналған құжаттардағы кеңістіктік-уақыттық формаларды анықтау автордың жеке өзі жасаған. Әр түрлі сөйлемдердің қайта жазылған нұсқаларының жиынтығын құру және олардың ұқсастығын бағалау әдісі де автордың жеке үлесі болып табылады. Алгоритм мен модельді эксперименттік бағалау ғылыми жетекшісі мен Ресей Ғылым академиясының Сібір бөлімшесі А. П. Ершов атындағы информатика жүйелері институтымен бірлесіп жүргізілді.

Қорғаудың негізгі ережелері - автордың жарияланған жұмысқа жеке үлесі.

Диссертация нәтижелерінің сыннан (апробация) өтуі. Диссертациялық жұмыста алынған нәтижелердің негізділігі дәлелденген зерттеу әдістерін қолдануға, жиын теориясы мен векторлық алгебраны, мәліметтерді өңдеудің статистикалық әдістерін дұрыс қолдануға негізделген, алынған нәтижелерді мәтіндерді табиғи тілде өңдеу және интеллектуалды шешімдерді қолдау саласындағы белгілі теориялық ережелер.

Ұсынылған әдістер мен алгоритмдердің сәйкестігі ұсынылған мәтіндік корпуста енгізілген интеллектуалды әдістердің нәтижелерімен, сондай-ақ апробация нәтижелерімен және интеллектуалды жүйенің прототипін енгізу актісімен расталады. Алынған нәтижелердің сенімділігі мәтіндерді кластерлеудің белгілі әдістері мен тәсілдерін, дұрыс статистикалық деректерді өңдеуді қолдану арқылы қамтамасыз етіледі.

Жұмыстың негізгі ережелері мен нәтижелері келесі ғылыми конференцияларда баяндалды және талқыланды:

- «Математиктер жүйесінде компьютерлік модельдеудің заманауи музделері Қазақстандық қайраткерлер индустриялық дамудың жадайында» Республикалық Ғылыми-практикалық конференция тезистер жинағы, (3-5 мамыр, 2018 ж., Астана қ.);

- “Gylym, bilim zhene Өндіріс интеграциясы - Ұлт жоспары іскер асырудың негізі” (№10 Сағынов ұлары) Халықтық Ғылым-практикалық конференциялар E N B E K T E R I (14-15 маусым 2018 ж.) 2 бөлім;

- «Ғылым мен техниканың жетістіктері» XIX Халықаралық ғылыми-практикалық конференция 15 наурыз 2019 ж. «Ғылыми баспа орталығы» өзектілігі. РФ », (15 наурыз, 2019 жыл, Мәскеу);

- BIG DATA және жоғары деңгейлі талдау, V халықаралық конференция материалдарының жинағы. (13-14 наурыз 2019 ж. Минск, Беларусь);

- BIG DATA және Advanced Analytics BIG DATA алтыншы халықаралық ғылыми-практикалық конференциясы және жоғары деңгейлі талдау (20-21 мамыр, 2020 ж., Минск, Беларусь);

- «GYLYM JANE BILIM - 2020» студенттер мен жас ғалымдардың XV Халықаралық ғылыми конференциясының материалдар жинағы, ҚР, (10.04.2020 ж., Нұр-Сұлтан).

Диссертациялық жұмыстың нәтижелері атындағы Ақпараттық жүйелер институтында енгізілді А.П. Ресей ғылым академиясының Сібір бөлімінің Ершовы, бұл ғылыми ережелерді енгізу және дипломдық жұмысты білім беру ұйымдарының тәжірибесіне енгізу туралы актілермен расталады.

Жарияланымдар. Диссертациялық зерттеудің нәтижелері бойынша 13 баспа жұмысы жарық көрді, оның ішінде 4 жұмыс ККСОН ұсынған рецензияланған баспа басылымдарында, 1 жұмыс Scopus индекстелген басылымдарда, және басқа ғылыми журналдарда 8 жарияланым және конференция материалдары. Авторлық құқығымен қорғалатын объектілерге құқықтардың мемлекеттік тізіліміне мәліметтерді енгізу туралы 1 авторлық куәлік алынды (ЭЕМ-ге арналған бағдарлама).

Диссертацияның құрылымы мен көлемі. Диссертациялық жұмыс ғылыми зерттеу құрылым логикасына сәйкес келеді, кіріспеден, үш бөлімнен, қорытындыдан және пайдаланылған әдебиеттер тізімі мен қосымшалардан тұрады. Жұмыстың көлемі - 115 бет, оның ішінде 22 сурет, 3 кесте, 3 қосымша. Әдебиеттер тізімі 107 атауды қамтиды.

Кіріспеде зерттеудің ғылыми аппараты келтіріліп, тақырыптың өзектілігі, оның теория мен практикадағы даму дәрежесі негізделеді, зерттеудің мақсаты, объектісі, пәні мен міндеттері айқындалады, ғылыми жаңалығы, теориялық және практикалық маңыздылығы ашылады, зерттеу әдістерін анықтайды, автордың жеке үлесін, жарияланымдар тізімін және жұмыс нәтижелерін апробациясы ұсынған.

1 бөлім семантикалық кітапханалардың логикалық-философиялық зерттеулерінің қазіргі жағдайына талдауды қамтиды. Бүгінгі таңда ғылыми және практикалық тұрғыдан өзекті болып табылатын белсенді дамып келе

жатқан ақпараттық іздеу әдістеріне талдау жасалды. Ақпаратты іздеудің келесі модельдері қарастырылған: буль моделі, векторлық модель, ықтималдық моделі. Кейбір модельдердің кемшіліктері ашылған. Қазіргі заманғы семантикалық кітапханаларды салыстырудың негізгі мәселелері анықталған.

2-бөлімде Link Grammar Parser және Диалинг анализаторларының табиғи тілдің ауызша сөйлемдерін формальды талдау сөздік мақалаларын және кеңістіктік-уақыттық қатынастарға қатысты көркем әдебиетте қолдану мысалдарын талдаулар зерттелген.

3-бөлім ағылшын, орыс және қазақ тілдеріндегі ақпаратты қамтитын семантикалық кітапхананың модельдері мен алгоритмдерін сипаттауға, сондай-ақ ұсыныстарды талдауды бағдарламалық іске асыруға арналған. Үш тілдің кеңістіктік-уақыттық конструкцияларын жіктеу үшін семантикалық кітапхана мен мәліметтер базасының моделін құруға арналған жүйенің интерфейсі сипатталған. Сондай-ақ үш тілде ақпараты бар семантикалық кітапхананың математикалық моделі келтірілген.

Қорытындыда зерттеу нәтижелері жинақталған, қорғауға ұсынылған ережелердің шынайылығын растайтын және дәлелдейтін негізгі қорытындылар келтірілген.

Қосымшада зерттеудің тәжірибелік материалдары ұсынылған.

Автор өзінің ғылыми жетекшісі, физика-математика ғылымдарының докторы, Ақпараттық жүйелер кафедрасының профессоры Түсіпов Джамалбек Алиаскарұлына диссертациялық зерттеу барысында проблеманы қойып, көмектескені үшін алғысын білдіреді.

Автор өзінің шетелдік консультанты, физика-математика ғылымдарының кандидаты, доцент, А.П. Ершов атындағы СБ РҒА Информатика жүйелері институты директорының ғылыми жұмыстар жөніндегі орынбасары Мурзин Федор Александровичке зерттеу кезінде кеңес бергені үшін алғысын білдіреді.

Сонымен қатар, автор Еуразия Ұлттық университеті Аударма теориясы мен практикасы кафедрасының профессоры, филология ғылымдарының докторы Тажибаева Сауле Жаксылықбаевнаға берген кеңестері үшін алғысын білдіреді.