

Жоба туралы қысқаша ақпарат

Жоба аты	AP19677835 «Қазақстан Республикасының заңнама саласындағы мемлекеттік тілге арналған семантикалық тәсілдерге негізделген интеллектуалды сұрақ-жауап үлгілерін зерттеу және жүйені әзірлеу» (0123PK00861)
Жоба өзектілігі	<p>Табиғи тілдерді өңдеу саласындағы заманауи жүйелер мен дамыған технологиялар - NLP (Natural language processing) және жасанды интеллект (AI) әртүрлі кәсіби қызмет салаларында және күнделікті адам өмірінде жетістіктерімен және даму және енгізу жылдамдығымен таң қалдырады. Сондай-ақ, NLP-де алынған нәтижелер мен жетістіктердің артында көптеген ғылыми жұмыстар мен зерттеулер, технологиялық шешімдер мен мүмкіндіктер жатқанын ескеру қажет. NLP-тің ең өзекті және танымал бағыттарының бірі - диалог және сұрақ-жауап жүйелері. Бұл жүйелерге көптеген қосалқы тапсырмалар кіреді, олардың нәтижелері жасанды интеллекттің әртүрлі салаларында қолданылуы мүмкін, мысалы, сөйлеу технологиялары, машиналық аударма, іздеу жүйелері және т.б.</p> <p>Сұрақ-жауап жүйесін шешудің әртүрлі ғылыми тәсілдері мен әдістері бар. Олардың кейбіреулері төменде ұсынылатын болады. Әртүрлі өндірушілер ұсынатын бағдарламалық қамтамасыз ету, мысалы, қазіргі уақытта сұрақ-жауап жүйелерін енгізудің өте көп саны бар. START сұрақ-жауап жүйесін енгізу назар аударуға тұрарлық. Ашық бастапқы коды мәртебесін талап ететін басқа қызықты жүйелер Cус, OpenEрhуgа және PIQUANT жүйелері болып табылады. Сондай-ақ, Semantic Web зерттеушілері OntoSearch, Semantic Wikis, Semantic Portals сияқты ағылшын тіліне семантикалық іздеу технологияларын, онтологияға негізделген маршруттардың (сұраулардың) көп агентті P2P семантикалық жүйесін, онтологияға негізделген сұрақ-жауап жүйелерін пайдалануды ұсынды. . Сондай-ақ әйгілі «Алиса» смарт станциясы - Яндекс жасаған виртуалды дауыстық көмекші. Табиғи сөйлеуді таниды, тірі диалогты имитациялайды, пайдаланушы сұрақтарына жауап береді және бағдарламаланған дағдылардың арқасында қолданбалы есептерді шешеді. Қазіргі таңда egov мемлекеттік порталында және түрлі банк жүйелерінде қазақ тілінде түрлі чатбот жүйелері қолданылады. Бірақ, өкінішке орай, олар тақырыппен және жауап хаттамаларымен шектелген, сонымен қатар дауыс функциясын қолдамайды.</p>

	<p>Өкінішке орай, қазақ тілі үшін бұл жүйелердің аналогтары (ашық қолжетімділік және/немесе ақылы ресурстар) жоқ. Жоғарыда аталған бағдарламалық өнімдер көптеген ресурстық (еуропалық, славян топтары) ағылшын, испан, орыс және т.б. тілдерге арналған. Өкінішке орай, түркі тілдері үшін (қазақ, қырғыз, түрік, өзбек және т.б.) қазіргі уақытта бар. қоғамдық доменді іске асыруда қолжетімді бағдарламалық құрал жоқ. Ұсынылған жұмыстар мен жүйелерді қазақ тіліне ішінара немесе толық қолдануға болмайды. Себебі, әрбір тіл үшін тілдік қасиеттерді ескере отырып, жеке қорлар жинақталып, өңделеді. Сондай-ақ әр тілдің өзіндік диалектісі, семантикалық (когнитивтік) ұғымдары, т.б.</p>
<p>Жоба мақсаты</p>	<p>Жобаның мақсаты – қазіргі семантикалық тәсілдер мен нейрондық желіні оқыту негізінде Қазақстан Республикасының заңнамасы саласында мәселені зерттеу және қазақ тілінде сұрақ-жауап технологиясын (алгоритмдер, модельдер, жүйелер) әзірлеу.</p>
<p>Жоба міндеттері</p>	<p>Жобаның негізгі міндеттері:</p> <ul style="list-style-type: none"> • Ашық электрондық көздерден мәтіндік деректерді жинау және өңдеу тәсілін әзірлеу; • Өзара әрекеттесу үшін сұрақ-жауап матрицасын құрастыру; • Машиналық оқыту негізінде семантиканы ескере отырып, қазақ тіліндегі жауапты синтездеу әдістемесін жасау. • Жіктеу моделін құру және нейрондық желіні оқыту арқылы жауаптарды болжау әдісін әзірлеу • Қазақ тіліндегі жүйенің дауыстық интерфейсін қолдау модулін әзірлеу және енгізу
<p>Күтілетін және қол жеткізілген нәтижелер</p>	<p>Жобадан күтілетін нәтижелер: мәселені зерттеу және қазақ тілінде диалог жүргізу мүмкіндігімен нейрондық желіні оқытуға негізделген интеллектуалды сұрақ-жауап технологиясының (алгоритмдер, модельдер, жүйелер) прототипін жасау. Web of Science дерекқорының Expanded Science Citation Index индексіне индекстелген жарияланымдарда мақалаларды жариялау және (немесе) Scopus деректер базасында CiteScore пайыздық көрсеткіші кемінде 35 (отыз бес) және мақаланың жариялануы KOKSNVO ұсынған рецензияланған шетелдік немесе отандық басылым.</p> <p>2023 жылға атқарылған жұмыстар және алынған нәтижелер:</p>

	<p>- Зерттеу тақырыбына аналитикалық шолу жасалды. Құқықтық және нормативтік құжаттарға қатысты сұрақ-жауап жүйелерінің, чат-боттардың және веб-порталдардың заманауи шетелдік аналогтары зерттеліп, сарапталды. ChatGPT және оның аналогтарының қасиеттері сипатталған, мысалы: «Alphachat», «Law ChatGPT», «DoNotPay», «LawDroid», «Ross Intelligence», «CaseMine». Олардың айырмашылықтары мен бақыланатын оқытудың қолданбалы әдістері мен шыңдалатын оқыту сипатталған.</p> <p>- Нормативтік құжаттардың түрлері зерттеліп, Қазақстан Республикасының әкімшілік-құқықтық актілерінің әзірленуіне және қабылдану тәртібіне қарай негізгі және туынды түрлерге жіктелді. ;- Қазақстан Республикасының заңнамалық құжаттарының иерархиясы сипатталған. Иерархия негізінде мәліметтер қорын қалыптастыру үшін заңнамалық құжаттардан деректер жиналады. – Қазақ тілінің заң шығару құжаттарының құрылымы формальды грамматика негізінде айқындалады. Осы мақсатта қазақ тіліндегі құқықтық құжаттардан түрлі үлгідегі мәтіндер мен сөйлемдерге талдау жасалды.</p> <p>- Мәліметтер Қазақстан Республикасының заңнамалық сайттарындағы мақалалардан жиналды. Негізгі ақпарат көзі – https://adilet.zan.kz/kaz/ сайты. Деректер html пішімінде жүктелді, содан кейін мәтіндік пішімге (txt) түрлендірілді. Деректерді жинау үшін біз Windows жүйесіне арналған әмбебап көп ағынды веб-сайтты талдау құралын қолдандық, ол көптеген функциялары және WEB мазмұнымен жұмыс істеудің барлық кезеңдерін толық басқару және конфигурациялау мүмкіндігі бар - Content downloader.</p>
<p>Зерттеу тобы мүшелерінің аты-жөні, идентификаторлары (Scopus Author ID, Researcher ID, ORCID, бар болса) және сәйкес профильдерге сілтемелер</p>	<p>Жоба жетекшісі: PhD. Рахимова Диана Рамазанқызы h-index: 4, Scopus авторының идентификаторы: 55682794500 Web of Science ResearcherID D-8421-2012</p> <p>Зерттеу тобының мүшелері</p> <ol style="list-style-type: none"> 1. Техника ғылымдарының докторы, проф. Тукеев Уалшер Әнуарбекұлы h-index: 5. Scopus авторының идентификаторы 55701639900 https://orcid.org/0000-0001-9878-981X 2. PhD Шормакова Әсем h-index:2, ORCID ID: 0000-0002-1637-4643, Author ID Scopus: 55786942400, Researcher ID Web of Science: D-5836-2015 3. PhD Қарибаева Айдана

	<p>h-индекс: 3 orcid идентификаторы: 0000-0002-2023-1573, Scopus идентификаторы: 57196004542, ResearcherID: AAR-4134-2020</p> <p>4. PhD Карюкин Владислав, h-1, https://orcid.org/0000-0002-8768-0349, Scopus Author ID: 57218952479 SciProfiles: 2410440</p> <p>5. Магистр Тұрарбек Әсем h-1, Scopus идентификаторы 57031944900, https://orcid.org/0000-0002-4793-0446</p> <p>6 Магистр Әмірова Дина h-индекс: 1, ResearcherID: P-5668-2017, Orcid идентификаторы: 0000-0002-0728-905X, Scopus ID: 57196009653</p>
Жарияланымдар тізімі (URL, DOI көрсетілген)	<p>2023 ж. жарияланымдар тізімі :</p> <p>1.Ualsher Tukeyev and etc. Kazakh-Tatar Machine Translation on the Base of Complete Set of Endings Model. Chapter of the book «Machine Learning». Editors: Burcu Arsan Ph.D(c) Prof. Dr. Natalya Ketenci ©Yeditepe University Press, pp. 73-90. ISBN: 978-975-307-139-0 Istanbul, 2023</p> <p>2.Diana Rakhimova, Yntymak Abdrazakh. Study and development of an approach for identifying incorrect words for the Kazakh language in semi-structured data. Chapter of the book «Machine Learning». Editors: Burcu Arsan Ph.D(c) Prof. Dr. Natalya Ketenci ©Yeditepe University Press, pp. 23-39. ISBN: 978-975-307-139-0 Istanbul, 2023.</p> <p>3.Тукеев У.А. Реляционные модели обработки тюркских языков. VIII — Международную научно-практическую конференцию «Информатика и прикладная математика» с 26 по 27 октября 2023 г. Алматы. 85-89 с.</p>
Патент туралы ақпарат	-