İTÜ
İSTANBUL TEKNİK ÜNİVERSİTESİ 1773

□ Zip Code : 34469
□ Tel. : +90-212-285-3601
□ Web : www.itu.edu.tr

REVIEW OF THE FOREIGN SCIENTIFIC CONSULTANT

on the dissertation work of **Asiya Kalisseitovna Sarsenbayeva** on the topic: "**Research of Models and Development of an Intelligent Question-and-Answer System Based on Semantic Approaches for the State Language in the Field of Legislation of the Republic of Kazakhstan**" submitted for the degree of Doctor of Philosophy (PhD) in the educational program: 8D06101 – Information Systems

The dissertation work of Asiya Kalisseitovna Sarsenbayeva is devoted to a significant and highly relevant scientific problem related to enhancing access to legislative information in the state language of the Republic of Kazakhstan. The rapid growth of artificial intelligence technologies and large language models has created new opportunities for the development of intelligent legal information systems; however, the adaptation of such technologies for low-resource and agglutinative languages, including Kazakh, remains insufficiently explored. In this regard, the topic of the dissertation is scientifically justified, timely, and of considerable practical significance for the development of digital legal technologies in Kazakhstan.

The dissertation demonstrates a comprehensive and systematic research approach. The author conducted an in-depth analysis of existing legal question-answering systems, semantic retrieval methods, and modern transformer-based language models. Special attention was paid to the linguistic and morphological characteristics of the Kazakh language and their impact on legal text processing and semantic understanding.

One of the important scientific contributions of the dissertation is the creation of a specialized legal dataset consisting of a 5,3 GB corpus of legislative documents and more than 8,000 question-answer pairs collected from official legal portals of the Republic of Kazakhstan. The preparation, preprocessing, annotation, and structuring of this dataset required substantial scientific and technical effort and provide a valuable contribution to the further development of Kazakh-language NLP resources.

The author proposed a novel two-stage semantic preprocessing approach named SemanticPrep, designed to transform natural language user queries into formalized legal representations and improve semantic retrieval quality. The proposed method combines mT5-based query formalization with semantic expansion techniques based on SVD/LSA models. This approach is particularly important for agglutinative languages where morphological variability significantly affects semantic matching accuracy.

Another important contribution of the dissertation is the integration of fine-tuned LLM models with Retrieval-Augmented Generation (RAG) architecture. The proposed pipeline combines semantic preprocessing, retrieval optimization, domain-adapted language generation, and legal semantic matching into a unified intelligent system. Experimental evaluations confirmed that the proposed architecture significantly improves the quality and relevance of generated legal responses compared to baseline approaches.

The dissertation contains extensive experimental studies and comparative analyses of baseline LLM models in both zero-shot and fine-tuned configurations. The obtained results demonstrate improvements in Recall@5, MRR, ROUGE-L, BERTScore, and Expert Score metrics, confirming the effectiveness of the proposed methods. The evaluation methodology is comprehensive and scientifically sound, combining both quantitative metrics and expert assessment procedures.

The dissertation is well-structured and logically organized. Theoretical foundations, methodological approaches, architectural models, mathematical formulations, implementation details, and experimental results are presented consistently and clearly. The presented figures, tables, and comparative analyses effectively support the scientific conclusions of the research.

It should also be emphasized that the research results were published in reputable international and national scientific journals and conference proceedings indexed in Scopus and Web of Science databases. This confirms both the scientific relevance of the research and the academic maturity of the doctoral candidate.

İTÜ

□ Zip Code : 34469
□ Tel. : +90-212-285-3601
□ Web : www.itu.edu.tr

The dissertation has clear theoretical and practical significance. The proposed intelligent legal question-answering system can be applied in digital government services, legal information systems, citizen support platforms, and educational resources aimed at improving access to legislative information in the Kazakh language. Furthermore, the developed approaches may serve as a foundation for future studies in legal NLP, semantic retrieval, and low-resource language technologies.

The main strengths of the dissertation include:

1. development of a specialized legal dataset for the Kazakh language;
2. proposal of a novel SemanticPrep semantic preprocessing method;
3. integration of LLM, semantic retrieval, and RAG technologies;
4. adaptation of transformer models for the legal domain in Kazakh;
5. comprehensive experimental validation and comparative analysis;
6. high scientific novelty and practical applicability.

In conclusion, I would like to note that the dissertation work of Asiya Kalisseitovna Sarsenbayeva represents a completed and independent scientific research study carried out at a high academic level. The dissertation fully satisfies the requirements for PhD dissertations in the field of Information Systems regarding scientific novelty, theoretical contribution, methodological rigor, and practical significance.

Therefore, **I strongly recommend** awarding Asiya Kalisseitovna Sarsenbayeva the degree of Doctor of Philosophy (PhD) in the educational program “8D06101 – Information Systems”.

Sincerely,

Eşref Adalı, Professor

Faculty of Computer and Informatics

Engineering Istanbul Technical

University Website:

http://web.itu.edu.tr/adali/

Email: adali@itu.edu.tr

Web: https://adalı.net

Phone: +90 532 314 2349

(лого: Стамбул техникалық университеті, 1773)
34469

Пошталық индекс:

Тел. : +90-212-285-3601
Веб : www.itu.edu.tr

8D06101-Ақпараттық Жүйелер білім беру бағдарламасы бойынша
Философия Докторы (PhD) ғылыми дәрежесін алу үшін ұсынылған "**Қазақстан Республикасының Заңнамасы Саласындағы Мемлекеттік Тілдің Семантикалық Тәсілдеріне Негізделген Интеллектуалды Сұрақ-Жауап Жүйесінің Модельдерін Зерттеу және Дамыту**" тақырыбындағы
Сәрсенбаева Әсия Қалисейітқызының диссертациялық жұмысына
ШЕТЕЛДІК ҒЫЛЫМИ КЕҢЕСШІНІҢ ПІКІРІ

Сәрсенбаева Әсия Қалисейітқызының диссертациялық жұмысы Қазақстан Республикасының мемлекеттік тіліндегі заңнамалық ақпаратқа қолжетімділікті арттыруға байланысты маңызды және өте өзекті ғылыми мәселеге арналған. Жасанды интеллект технологиялары мен ірі тілдік модельдердің қарқынды өсуі интеллектуалды құқықтық ақпараттық жүйелерді дамытуға жаңа мүмкіндіктер туғызды; дегенмен, мұндай технологиялардың ресурстары аз және агглютинативті тілдерге, оның ішінде қазақ тіліне бейімделуі әлі де жеткілікті түрде зерттелмеген. Осыған байланысты диссертация тақырыбы ғылыми негізделген, уақтылы және Қазақстанда цифрлық құқықтық технологияларды дамыту үшін айтарлықтай практикалық маңызы бар.

Диссертация жан-жақты және жүйелі зерттеу тәсілін көрсетеді. Автор қолданыстағы құқықтық сұрақтарға жауап беру жүйелеріне, семантикалық іздеу әдістеріне, трансформаторға негізделген заманауи тілдік модельдерге терең талдау жасады. Қазақ тілінің лингвистикалық және морфологиялық ерекшеліктеріне және олардың мәтінді құқықтық өңдеуге және семантикалық түсінуге әсеріне ерекше назар аударылды.

Диссертацияның маңызды ғылыми үлестерінің бірі-5,3 ГБ заңнамалық құжаттар корпусынан және Қазақстан Республикасының ресми құқықтық порталдарынан жиналған 8000-нан астам сұрақ-жауап жұптарынан тұратын мамандандырылған құқықтық мәліметтер базасын құру. Бұл деректер жинағын дайындау, алдын ала өңдеу, аңдатпалау және құрылымдау айтарлықтай ғылыми-техникалық күш-жігерді қажет етті және қазақ тіліндегі Табиғи Тілді Өңдеу (NLP) ресурстарын одан әрі дамытуға құнды үлес қосуды қамтамасыз етті.

Автор табиғи тілді қолданушылардың сұраныстарын формальды заңды өкілдіктерге айналдыруға және семантикалық іздеу сапасын жақсартуға арналған SemanticPrep деп аталатын жаңа екі сатылы семантикалық алдын-ала өңдеу тәсілін ұсынды. Ұсынылған әдіс mT5 негізіндегі сұрауды ресімдеуді Матрицаның Сингулярлық Ыдырауы (SVD) / Жасырын Семантикалық Талдау (LSA) үлгілеріне негізделген семантикалық кеңейту әдістерімен біріктіреді. Бұл тәсіл морфологиялық өзгергіштік семантикалық сәйкестіктің дәлдігіне айтарлықтай әсер ететін агглютинативті тілдер үшін өте маңызды.

Диссертацияның тағы бір маңызды үлесі-дәл бапталған Үлкен Тілдік Модельдер (LLM) модельдерін Кеңейтілген Іздеу Генерациясының (RAG) архитектурасымен біріктіру. Ұсынылған құбыр семантикалық алдын-ала өңдеуді, іздеуді оңтайландыруды, доменге бейімделген тіл генерациясын және заңды семантикалық сәйкестікті бірыңғай интеллектуалды жүйеге біріктіреді. Эксперименттік бағалаулар ұсынылған архитектураның негізгі тәсілдермен салыстырғанда жасалған құқықтық жауаптардың сапасы мен өзектілігін айтарлықтай жақсартатынын растады.

Мекен-жайы: Маслак ауданы, Решат Нури көшесі, №: 13, 34467 Сарыер/Стамбул, Түркия

(лого: Стамбул техникалық университеті, 1773)
34469

Пошталық индекс:

Тел. : +90-212-285-3601
Веб : www.itu.edu.tr

Диссертацияда ауқымды эксперименттік зерттеулер мен LLM базалық модельдерінің нөлдік және дәл бапталған конфигурациялардағы салыстырмалы талдаулары бар. Алынған нәтижелер ұсынылған әдістердің тиімділігін растайтын Recall@5, MRR, ROUGE-L, BERTScore және Сараптамалық Бағалау көрсеткіштерінің жақсарғанын көрсетеді. Бағалау әдістемесі сандық көрсеткіштерді де, сараптамалық бағалау процедураларын да біріктіретін жан-жақты және ғылыми негізделген.

Диссертация жақсы құрылымдалған және логикалық түрде ұйымдастырылған. Теориялық негіздер, әдіснамалық тәсілдер, архитектуралық модельдер, математикалық тұжырымдар, іске асырудың егжей-тегжейлері, эксперимент нәтижелері дәйекті және нақты ұсынылған. Ұсынылған суреттер, кестелер және салыстырмалы талдаулар зерттеудің ғылыми қорытындыларын тиімді түрде негіздейді.

Сондай-ақ, зерттеу нәтижелері беделді халықаралық және ұлттық ғылыми журналдарда және Scopus және Web Of Science дерекқорларында индекстелген конференция материалдарында жарияланғанын атап өткен жөн. Бұл зерттеудің ғылыми өзектілігін де, докторанттың академиялық жетілуін де растайды.

Диссертацияның нақты теориялық және практикалық маңызы бар. Ұсынылған интеллектуалды құқықтық сұрақтарға жауап беру жүйесі цифрлық мемлекеттік қызметтерде, құқықтық ақпараттық жүйелерде, азаматтарды қолдау платформаларында және қазақ тіліндегі заңнамалық ақпаратқа қолжетімділікті жақсартуға бағытталған білім беру ресурстарында қолданылуы мүмкін. Сонымен қатар, әзірленген тәсілдер құқықтық NLP, семантикалық іздеу және ресурстары аз тілдік технологиялар саласындағы болашақ зерттеулердің негізі бола алады.

Диссертацияның негізгі күшті жақтарына мыналар жатады:

1. қазақ тілі бойынша мамандандырылған құқықтық деректер қорын әзірлеу;
2. жаңа SemanticPrep семантикалық алдын ала өңдеу әдісінің ұсынысы;
3. LLM, семантикалық іздеу және RAG технологияларын интеграциялау;
4. трансформаторлық модельдерді құқықтық салаға қазақ тілінде бейімдеу;
5. кешенді эксперименттік валидация және салыстырмалы талдау;
6. жоғары ғылыми жаңалығы және практикалық қолданылуы.

Қорытындылай келе, Сәрсенбаева Әсия Қалисейітқызының диссертациялық жұмысы жоғары академиялық деңгейде жүргізілген, аяқталған және тәуелсіз ғылыми-зерттеу жұмыстары екенін атап өткім келеді. Диссертация ғылыми жаңалыққа, теориялық үлеске, әдістемелік қатаңдыққа, практикалық маңыздылыққа қатысты Ақпараттық Жүйелер саласындағы кандидаттық диссертацияларға қойылатын талаптарды толығымен қанағаттандырады.

Сондықтан, **мен** Сәрсенбаева Әсия Қалисейітқызына "8D06101 – Ақпараттық Жүйелер" білім беру бағдарламасы бойынша Философия Докторы (PhD) дәрежесін **беруді ұсынамын**.

Құрметпен,
/қол қойылған/
Эшреф Адалы, Профессор
Стамбул Техникалық Университетінің
Компьютерлік және Компьютерлік инженерия факультеті
Университеттің веб-сайты: http://web.itu.edu.tr/adali/

Мекен-жайы: Маслак ауданы, Решат Нури көшесі, №: 13, 34467 Сарыер/Стамбул, Түркия

(лого: Стамбул техникалық университеті, 1773)
34469

Пошталық индекс:

Тел. : +90-212-285-3601
Веб : www.itu.edu.tr

Электрондық пошта: adali@itu.edu.tr
Веб: https://adalı.net
Телефон: +90 532 314 2349

Мекен-жайы: Маслак ауданы, Решат Нури көшесі, №: 13, 34467 Сарыер/Стамбул, Түркия