

## Жоба туралы қысқаша ақпарат

Жоба аты	AP09261344 «Географиялық ақпараттық жүйелерді ақпараттық қамтамасыз ету үшін гетерогенді көздерден геокеңістіктік объектілерді автоматты түрде алу әдістерін әзірлеу» (0121PK00388)
Жоба өзектілігі	Жұмыстың өзектілігі әртүрлі ашық ақпарат көздерінен, әсіресе мәтіндік веб-кестелерден алынған геокеңістіктік деректерді тиімді өңдеу және біріктіру қажеттілігінде жатыр. Форматтардың үйлесімсіздігі мен семантиканың болмауына байланысты қиындықтар интеграция процесін қиындатады және шешім қабылдау үшін негізгі ақпараттың жетіспеуіне әкелуі мүмкін. Мәтіндік веб-кестелерден геокеңістіктік объектілерді және олардың атрибуттарын автоматты түрде алу әдістерін әзірлеу геоақпараттық жүйелерді тиімдірек пайдалануды және географиялық контексте негізделген шешімдер қабылдау үшін деректерді талдау сапасын жақсартуды қамтамасыз ететін осы процесті жеңілдету жолындағы маңызды қадам болып табылады.
Жоба мақсаты	Жобаның мақсаты гетерогенді ашық деректер көздерінен, атап айтқанда мәтіндік веб-кестелерден геокеңістіктік объектілерді және олармен байланысты кеңістіктік емес атрибуттарды автоматты түрде алу әдістерін әзірлеу болып табылады.
Жоба міндеттері	Жобаның мақсатына жету үшін келесі міндеттерді шешу қажет: 1. Атрибут-мән жұптарының жиынтығы ретінде мәтіндік веб-кестелерден геокеңістіктік объектілерді және олардың сандық және сапалық сипаттамаларын алу мен түсіндірудің семантикалық әдістерін зерттеу және әзірлеу. 2. Семантикалық Вебтің ашық геоақпараттық ресурстарына негізделген алынған геокеңістіктік деректерді біріктіру және теңестіру әдістерін зерттеу және әзірлеу. 3. Құрылған әдістер мен алгоритмдерді біртұтас технологияға біріктіру, оның негізін схемасыз таратылған NoSQL моделі құрайды. 4. Әзірленген технология негізінде бағдарламалық өнімді прототиптеу. "Туризм", "Төтенше жағдайлар" домендеріндегі веб-сайттардан геокеңістіктік ақпаратты талдауға және алуға арналған веб-сервистерді құру
Күтілетін және қол жеткізілген нәтижелер	Жобаның қорытындысы бойынша: – Web of Science дерекқорының Expanded Science Citation Index индексында және Scopus дерекқорында CiteScore процентілі кемінде 50 (елу) болатын кемінде 3 (үш) мақала мен (немесе) шолулар рецензияланған ғылыми басылымдарда жарияланады; – және Scopus деректер базасында индекстелген халықаралық конференциялар материалдарындағы 2 мақала, мысалы Computational Collective Intelligence Conference; – ҚР КОКСОН ұсынған рецензияланған шетелдік немесе отандық басылымда кемінде 3 (үш) мақала немесе шолу;

	<p>– және Қазақ баспасында 1 монография (Қазақ университеті);</p> <p>– авторлық құқық объектісіне құқықтарды мемлекеттік тіркеу туралы авторлық куәлік алынады.</p> <p>Жобаны аяқтау, геокеңістіктік объектілерді автоматты түрде алу технологиясын сәтті пайдалану үшін бағдарламалық технологияларды сынақтан өткізу нәтижесінде ғылыми-техникалық құжаттаманы әзірлеу жоспарлануда.</p> <p>Қол жеткізілген нәтижелер:</p> <ul style="list-style-type: none"> <li>- атрибут-мән жұптарының жиындары ретінде мәтіндік кестелерден деректерді алудың интеллектуалды әдістері, веб-кестелердің физикалық, функционалдық және логикалық құрылымын талдау әдістері және кіріс деректер түріне байланысты веб-кестелерді тану үшін сәйкес талдаушылар әзірленді.</li> <li>- Геодеректерді семантикалық интерпретациялау әдістері әзірленді, оның ішінде құрылымдалмаған кілттік-мәнді жадқа деректерді үлестірілген жүктеу, онтологиялық тәсіл негізінде деректерді объектілік бейнелеуге семантикалық түрлендіру, алынған геодеректердің координаталық анықтамасын анықтау және нақтылау; деректер.</li> <li>- Интернеттегі мәтіндік кестелерден геоақпаратты автоматты түрде алу технологиясы әзірленді, бұлтқа негізделген таратылған инфрақұрылым, құрылған әдістер мен алгоритмдерді бір қызметке біріктіру.</li> </ul>
<p>Зерттеу тобы мүшелерінің аты-жөні, идентификаторлары (Scopus Author ID, Researcher ID, ORCID, бар болса) және сәйкес профильдерге сілтемелер</p>	<ol style="list-style-type: none"> <li>1. Мансурова Мадина Есимхановна - физика-математика ғылымдарының кандидаты, доцент, әл-Фараби атындағы ҚазҰУ «Жасанды интеллект және Big Data» кафедрасының меңгерушісі, әл –Фараби атындағы ҚазҰУ жетекші ғылыми қызметкері. Scopus H-index =5, Web of Science H-index = 2, Scopus-те индекстелген жарияланымдар – 64, дәйексөздердің жалпы саны – 91.</li> <li>2. Нугуманова Алия Багдатовна – PhD, Big Data and Blockchain Technologies Astana IT University ғылыми-зерттеу орталығының директоры. Scopus Author ID: 55864815200, Orcid ID: 0000-0001-5522-4421, h-index=5.</li> <li>3. Барахнин Владимир Борисович –жоғары білімі бар, Новосібір мемлекеттік университетін бітірді, ResearchGate: A-5856-2014, ORCID: <a href="https://orcid.org/0000-0003-3299-0507">https://orcid.org/0000-0003-3299-0507</a>, SCOPUS: 6508258628.</li> <li>4. Шоманов Адай Сакенович –PhD докторы, Назарбаев Университетінің қызметкері, Scopus Author ID: 57195543732, h-index Scopus = 4.</li> <li>5. Оспан Әсел Ғалымжанқызы – магистр, ақпараттық технологиялар факультетінің аға оқытушысы. Scopus Author ID: 57238489800, ORCID ID: 0000-0002-1860-6997, h-index=1.</li> </ol>
<p>Жарияланымдар тізімі (URL, DOI көрсетілген)</p>	<ol style="list-style-type: none"> <li>1. Mansurova M, Barakhnin V, Ospan A, Titkov R. Ontology-Driven Semantic Analysis of Tabular Data: An Iterative Approach with Advanced Entity Recognition. Appl Sci. 2023;13(19):10918. doi:10.3390/app131910918.</li> </ol>

2. Kadyrbek N, Mansurova M, Shomanov A, Makharova G. The Development of a Kazakh Speech Recognition Model Using a Convolutional Neural Network with Fixed Character Level Filters. *Big Data Cogn Comput.* 2023;7(3):132. doi:10.3390/bdcc7030132.
3. Ospan A, Mansurova M, Barakhnin V, Nugumanova A, Titkov R. The Development of a Water Resource Monitoring Ontology as a Research Tool for Sustainable Regional Development. *Data.* 2023;8(11):162. doi:10.3390/data8110162.
1. K. Bauyrzhan, M. Madina and O. Assel, "Fine-Tuning the Wav2vec2 Model for Kazakh Speech: A Study on a Limited Corpus," *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)*, Astana, Kazakhstan, 2023, pp. 124-128, doi: 10.1109/SIST58284.2023.10223504.
2. Barakhnin V, Mansurova M, Grigorieva I, Kozhemyakina O, Ospan A. TableProcessor: The Tool for the Analysis and the Interpretation of Web Tables to Create the Geo Knowledge Base of Kazakhstan. In: Dolinina O, et al., eds. *Artificial Intelligence in Models, Methods and Applications. AIES 2022. Studies in Systems, Decision and Control*, vol 457. Springer; 2023:219-229. doi:10.1007/978-3-031-22938-1\_15. Accessed April 25, 2023.
3. Mansurova M, Ospan A, Kakimzhanov Y, Resnik B, Tyulyubayev D. Development of an Application for Monitoring and Analyzing the Dynamics of the Tuyuk Su Mountain Glacier. *SIST 2022 International Conference on Smart Information Systems and Technologies*. <https://sist.astanait.edu.kz/wp-content/uploads/2022/05/conference-programme-129.pdf>. Published 2022.
4. Mansurova M, Barakhnin V, Kyrgyzbayeva M, Kadyrbek N. Named Entity Extraction Model Based on the Random Walk Method. In: *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*. IEEE; 2021. <https://ieeexplore.ieee.org/document/9465992>.
5. Ospan A, Mansurova M, Kakimzhanov E, Aldakulov B. KazRivDyn: Toolkit for Measuring the Dynamics of Kazakhstan Rivers with Graphics Based on Google Earth Engine. In: *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*. IEEE; 2021. <https://ieeexplore.ieee.org/document/9465902>.
6. Akhmed-Zaki D, Mansurova M, Yertuyak A, Chikibayeva D. Development of Web Application for Visualizing City Emergencies. *2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)*. IEEE; 2021. doi:10.1109/SIST50301.2021.9465919.
7. Meiran Zhiyenbayev, Assel Ospan, Nadezhda Kunicina, Madina Mansurova, Roman Titkov. Systematic data procurement in an owl-embedded information and analytical framework for the monitoring of water resources in the Ile-Balkhash basin. *Scientific Journal of Astana IT University*, ISSN

	<p>(P): 2707-9031 ISSN (E): 2707-904X, Volume 15, September 2023.</p> <p>8. Zhiyenbayev M, Ospan A, Mansurova M. ETL Process for Water Resources and Demographics Data: An Open Source Data Processing Tools and Visualizations. Vestn Nats Inzh Akad Respub Kaz. 2023;(88):38-48. doi:10.47533/2023.1606-146X.4.</p> <p>12. Nugumanova A, Apayev K, Baiburin Y, Mansurova M, Ospan A. QURMA: A Table Extraction Pipeline for Knowledge Base Population. J Math Mech Comput Sci. 2022;114(2). <a href="https://bm.kaznu.kz/index.php/kaznu/article/view/1086">https://bm.kaznu.kz/index.php/kaznu/article/view/1086</a>. Published June 2022. Accessed October 19, 2022. doi:10.26577/JMMCS.2022.v114.i2.08.</p> <p>13. Ospan A, Mansurov M, Kakimzhanov E, Ixanov S, Barakhnin V. Development of a Program for the Integration of Socio-Economic Indicators with Spatial Data to Analyze the Standard of Living of the Population of Kazakhstan. Vestn Nats Inzh Akad Respub Kaz. 2022;(85):67-78. doi:10.47533/2020.1606-146X.170.</p> <p><b>Монография:</b></p> <p>1. М.Е. Мансурова. Text Mining озық модельдері мен әдістері: монография. – Алматы: Қазақ университеті, 2023. – 112 с. ISBN 978-601-04-6499-5</p>
<p>Патент туралы ақпарат</p>	<p><b>Авторлық куәліктер:</b></p> <p>1. Оспан Әсел, Мансурова Мадина Есимхановна. Білім графиктерін толықтыру үшін гетерогенді көздерден алынған кестелерді семантикалық талдауға арналған итеративті алгоритм. 2023 жылғы "28" қыркүйектегі № 39296 ЭЕМ үшін бағдарламаны мемлекеттік тіркеу туралы куәлік.</p> <p>2. Мансурова Мадина Есимхановна, Қадырбек Нұрғали, Досанов Бекжан, Қырғызбаева Маржан, Тюлепбердинова Гульнур. Қазақ тіліндегі мәтіндерді алдын ала өңдеу конвейері. 2021 жылғы "21" мамырдағы № 17792.</p> <p>3. BI-LSTM негізінде қазақ тіліндегі жаңалықтар көздерінен аталған нысандарды алу алгоритмі. 2021 жылғы "12" мамырдағы № 17402.</p>
<p>Бейнематериал: <a href="https://youtu.be/CF0ie1zDX1E">https://youtu.be/CF0ie1zDX1E</a></p>	