

Краткая информация о проекте

Наименование	AP09261344 Разработка методов автоматического извлечения геопространственных объектов из гетерогенных источников для информационного обеспечения геоинформационных систем. (0123РК00131)
Актуальность	Актуальность работы заключается в неотъемлемой потребности в эффективной обработке и интеграции геопространственных данных из различных открытых источников, особенно из текстовых веб-таблиц. Сложности, связанные с несовместимостью форматов и отсутствием семантики, затрудняют процесс интеграции и могут привести к упущению ключевой информации для принятия решений. Разработка методов автоматического извлечения геопространственных объектов и их атрибутов из текстовых веб-таблиц представляет собой важный шаг в направлении упрощения этого процесса, обеспечивая более эффективное использование геоинформационных систем и повышение качества анализа данных для принятия обоснованных решений в географическом контексте.
Цель	Целью проекта является разработка методов автоматического извлечения геопространственных объектов и связанных с ними непространственных атрибутов из гетерогенных открытых источников данных, а именно из текстовых веб-таблиц.
Задачи	Для достижения цели проекта нужно решить следующие задачи: 1. Исследование и разработка семантических методов извлечения и интерпретации геопространственных объектов и их количественных и качественных описаний из текстовых веб-таблиц как наборов пар «атрибут-значение». 2. Исследование и разработка методов интеграции и выравнивания извлеченных геопространственных данных на основе открытых геоинформационных ресурсов Семантического Веба. 3. Консолидация созданных методов и алгоритмов в единую технологию, основу которой образует бессхемная распределенная NoSQL модель. 4. Прототипирование программного продукта на основе разработанной технологии. Создание веб-сервисов для парсинга и извлечения геопространственной информации с веб-сайтов в доменах «Туризм», «Чрезвычайные ситуации».
Ожидаемые и достигнутые результаты	По результатам проекта: – будут опубликованы не менее 3 (трех) статей и (или) обзоров в рецензируемых научных изданиях, индексируемых в Science Citation Index Expanded базы Web of Science и (или) имеющих процентиль по CiteScore в базе Scopus не менее 50 (пятидесяти); – и 2 статьи в трудах международных конференций, индексируемых в базе данных Scopus, например, Computational Collective Intelligence Conference;

	<p>– не менее 3 (трех) статей или обзоров в рецензируемом зарубежном или отечественном издании, рекомендованном КОКСОН РК;</p> <p>– и 1 монография в казахстанском издательстве (Қазақ университеті);</p> <p>– будет получено авторское свидетельство о государственной регистрации прав на объект авторского права.</p> <p>В результате завершения проекта, проверки программных технологий для успешного использования технологии автоматического извлечения геопространственных объектов планируется разработка научно-технической документации.</p> <p>Достигнутые результаты:</p> <ul style="list-style-type: none"> - Разработаны интеллектуальные методы извлечения данных из текстовых таблиц как наборов пар “атрибут-значение”, методы анализа физической, функциональной и логической структуры веб-таблиц и соответствующие парсеры для распознавания веб-таблиц в зависимости от типа входных данных. - Разработаны методы семантической интерпретации геоданных, включающей в себя распределенную загрузку данных в неструктурированное хранилище “ключ-значение”, семантическую трансформацию данных в объектное представление на основе онтологического подхода, определение и уточнение координатной привязки извлеченных геоданных с помощью извлеченных данных. - Разработана технология автоматического извлечения геоинформации из текстовых таблиц Веба, облачная распределенная инфраструктура с консолидацией созданных методов и алгоритмов в единый сервис.
<p>Имена и фамилии членов исследовательской группы с их идентификаторами (Scopus Author ID, Researcher ID, ORCID, при наличии) и ссылками на соответствующие профили</p>	<ol style="list-style-type: none"> 1. Мансурова Мадина Есимхановна - Кандидат физико-математических наук, доцент, заведующая кафедрой искусственного интеллекта и Big Data КазНУ им. аль-Фараби, ведущий научный сотрудник КазНУ им. аль-Фараби. Scopus H-index =5, Web of Science H-index = 2, публикаций, индексируемых в Scopus – 64, общее количество цитирований – 91. 2. Нугуманова Алия Багдатовна – PhD, директор НИЦ Big Data and Blockchain Technologies Astana IT University. Scopus Author ID: 55864815200, Orcid ID: 0000-0001-5522-4421, h-index=5. 3. Барахнин Владимир Борисович – образование высшее, окончил Новосибирский госуниверситет, ResearchGate: A-5856-2014, ORCID: https://orcid.org/0000-0003-3299-0507, SCOPUS: 6508258628.

	<p>4. Шоманов Адай Сакенович – доктор PhD, сотрудник Назарбаев Университета, Scopus Author ID: 57195543732, h-index Scopus = 4.</p> <p>5. Оспан Әсел Ғалымжанқызы – магистар, старший преподаватель факультета информационных технологий. Scopus Author ID: 57238489800, ORCID ID: 0000-0002-1860-6997, h-index=1.</p>
<p>Список публикаций со ссылками на них</p>	<p>1. Mansurova M, Barakhnin V, Ospan A, Titkov R. Ontology-Driven Semantic Analysis of Tabular Data: An Iterative Approach with Advanced Entity Recognition. <i>Appl Sci.</i> 2023;13(19):10918. doi:10.3390/app131910918.</p> <p>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. <i>Big Data Cogn Comput.</i> 2023;7(3):132. doi:10.3390/bdcc7030132.</p> <p>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. <i>Data.</i> 2023;8(11):162. doi:10.3390/data8110162.</p> <p>4. K. Bauyrzhan, M. Madina and O. Assel, "Fine-Tuning the Wav2vec2 Model for Kazakh Speech: A Study on a Limited Corpus," <i>2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)</i>, Astana, Kazakhstan, 2023, pp. 124-128, doi: 10.1109/SIST58284.2023.10223504.</p> <p>5. 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. <i>Artificial Intelligence in Models, Methods and Applications. AIES 2022. Studies in Systems, Decision and Control</i>, vol 457. Springer; 2023:219-229. doi:10.1007/978-3-031-22938-1_15. Accessed April 25, 2023.</p> <p>6. 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. <i>SIST 2022 International Conference on Smart Information Systems and Technologies</i>. https://sist.astanait.edu.kz/wp-content/uploads/2022/05/conference-programme-129.pdf. Published 2022.</p> <p>7. Mansurova M, Barakhnin V, Kyrgyzbayeva M, Kadyrbek N. Named Entity Extraction Model Based on the Random Walk Method. In: <i>2021 IEEE International Conference on Smart Information Systems and Technologies (SIST)</i>. IEEE; 2021. https://ieeexplore.ieee.org/document/9465992.</p> <p>8. 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: <i>2021 IEEE International Conference on Smart</i></p>

	<p>Information Systems and Technologies (SIST). IEEE; 2021. https://ieeexplore.ieee.org/document/9465902.</p> <p>9. 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.</p> <p>10. 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): 2707-9031 ISSN (E): 2707-904X, Volume 15, September 2023.</p> <p>11. 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). https://bm.kaznu.kz/index.php/kaznu/article/view/1086. 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 style="text-align: center;">Монография:</p> <p>1. М.Е. Мансурова. Передовые модели и методы Text Mining: монография. – Алматы: Қазақ университеті, 2023. – 112 с. ISBN 978-601-04-6499-5</p>
Информация о патентах	<p>Авторские свидетельства:</p> <p>1. Оспан Әсел, Мансурова Мадина Есимхановна. Итеративный алгоритм для семантического анализа таблиц из гетерогенных источников для пополнения графов знаний. Свидетельство о государственной регистрации программы для ЭВМ № 39296 от «28» сентября 2023 года.</p> <p>2. Мансурова Мадина Есимхановна, Қадырбек Нұрғали, Досанов Бекжан, Қырғызбаева Маржан, Түлепбердинова Гүльнур. Конвейер предварительной обработки текстов на казахском языке. № 17792 от «21» мая 2021 года.</p> <p>3. Мансурова Мадина Есимхановна, Чикибаева Дарья Юрьевна, Түлепбердинова Гүльнур Алпыскызы.</p>

	Алгоритм извлечения именованных сущностей из новостных источников на казахском языке на основе bi-LSTM. № 17402 от «12» мая 2021 года.
--	--

Видео: https://youtu.be/CF0ie1zDX1E
--