doi 10.18699/bgrs2026-358

Автоматическая реконструкция и анализ генных сетей в области агробиологии и биомедицины с использованием когнитивной системы ANDSystem

Иванисенко В.А.*, Деменков П.С., Иванисенко Т.В., Клещев М.А., Антропова Е.А., Яцык И.В., Волянская А.Р., Адамовская А.В.
[1] *Институт цитологии и генетики СО РАН, Новосибирск, Россия*
** salix@bionet.nsc.ru*

Ключевые слова: искусственный интеллект; когнитивная платформа; генные сети; ANDSystem; SmartCrop; биомедицина; агробиология; омиксные данные; стрессовый ответ; лекарственные мишени; биомаркеры

Мотивация и цель: Быстрое развитие высокопроизводительных технологий в геномике, транскриптомике, протеомике и метаболомике в биологии растений, биомедицине и смежных областях привело к накоплению огромных массивов генетических и молекулярно-биологических данных, распределённых по тысячам фактографических баз данных. Методы реконструкции и анализа генных сетей формируют основу для интеграции и системного анализа омиксных данных, обеспечивая их интерпретацию за счёт выявления причинно-следственных связей и реконструкции молекулярно-генетических механизмов, лежащих в основе фенотипических признаков и биологических процессов на клеточном и организменном уровнях. Вместе с тем эффективная реконструкция генных сетей требует разработки вычислительных методов автоматического извлечения знаний из десятков миллионов научных публикаций с использованием методов машинного обучения и искусственного интеллекта.
Целью данной работы являлось представление результатов разработки когнитивной системы ANDSystem, обеспечивающей полный цикл инженерии знаний – от автоматического анализа научных публикаций до реконструкции генных сетей.
Методы и алгоритмы: Автоматическое извлечение знаний из естественно-языковых текстов основывалось на комплексном подходе, интегрирующем онтологии предметной области, методы машинного обучения для формирования семантико-лингвистических правил, обеспечивающих быстрый анализ текстов, а также методы искусственного интеллекта, включая графовые нейронные сети и большие языковые модели. Реконструкция генных сетей осуществлялась с использованием оригинальных подходов, основанных на онтологическом описании предметной области, включающем понятия и типы взаимоотношений между биологическими сущностями.
Результаты: ANDSystem представляет собой когнитивную программно-информационную систему, предназначенную для работы с множеством типов биологических объектов и автоматического извлечения знаний о различных типах взаимодействий из научных текстов [1–3]. База знаний ANDSystem, сформированная на основе автоматического анализа аннотаций PubMed, полнотекстовых статей PMC и фактографических баз данных, содержит формализованные сведения о более чем 150 миллионах взаимодействий между молекулярно-генетическими объектами различных организмов, превосходя сопоставимые мировые системы по числу поддерживаемых типов объектов и взаимоотношений.
В агробиологии ANDSystem послужила основой для разработки системы SmartCrop, предназначенной для изучения молекулярно-генетических механизмов, определяющих связи между генотипом и фенотипом у сельскохозяйственно значимых культур риса и пшеницы [4]. В частности, SmartCrop применялась для идентификации регуляторных и сигнальных путей, участвующих в ответе растений на специфические условия роста, а также на биотические и абиотические стрессовые воздействия, включая высокую и низкую температуру, засуху, засоление, загрязнение почв металлами, воздействие удобрений, гормонов и элиситоров. Кроме того, система использовалась для предсказания генов-кандидатов для генотипирования, поиска маркеров для маркер-ассоциированной селекции и идентификации молекулярных мишеней для направленного воздействия различных веществ и внешних факторов.
В биомедицине ANDSystem применялась для интерпретации омиксных данных, реконструкции молекулярных механизмов заболеваний, анализа механизмов лекарственно-лекарственных взаимодействий, исследования токсических эффектов химических соединений, идентификации лекарственных мишеней, поиска новых биомаркеров и выявления генетических факторов риска заболеваний. В частности, система использовалась для изучения молекулярно-генетических механизмов взаимодействия патоген–хозяин, реконструкции сигнальных путей клеточных процессов, регулируемых вирусными белками, приоритизации фармакологических мишеней для терапии рака печени, исследования регуляторной роли неструктурных белков SARS-CoV-2 в развитии метаболических нарушений у пациентов с COVID-19, а также для анализа данных метаболомного скрининга плазмы крови пациентов с послеоперационным делирием с целью реконструкции молекулярно-генетических механизмов нарушений метаболизма сфинголипидов..
Заключение: Таким образом, ANDSystem представляет собой мощную когнитивную систему для

реконструкции и анализа генных сетей, обладающую уникальной базой знаний, превосходящей известные аналоги по разнообразию представленных биологических объектов и типов взаимодействий. Система успешно применяется для решения широкого круга задач в области агробиологии и биомедицины.
Финансирование: The bioinformatics analysis was supported by the budget project, FWNR-2022-0020.
Благодарности: Авторы выражают благодарность Гайслеру Е.В., Рогачеву А.Д., Басову Н.В., Покровскому А.Г. за получение и анализ метаболомных данных.

Automated reconstruction and analysis of gene networks in the field of agrobiology and biomedicine using cognitive system ANDSystem

Ivanisenko V.A.*, Demenkov P.S., Ivanisenko T.V., Kleshchev M.A., Antropova E.A., Yatsyk I.V., Volyanskaya A.R., Adamovskaya A.V.
[1] *Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia*
** salix@bionet.nsc.ru*

Key words: artificial intelligence; cognitive platform; gene networks; ANDSystem; SmartCrop; biomedicine; agrobiology; omics data; stress response; drug targets; biomarkers

Motivation and Aim: The rapid development of high-throughput experimental technologies in genomics, transcriptomics, proteomics, and metabolomics across plant biology, biomedicine, and related fields has led to the accumulation of vast amounts of genetic and molecular biological data distributed across thousands of factual databases. Methods for the reconstruction and analysis of gene networks provide a platform for the integration and systems-level analysis of omics data, enabling their interpretation through the identification of causal relationships and the reconstruction of molecular genetic mechanisms underlying phenotypic traits and biological processes at the cellular and organismal levels. At the same time, effective reconstruction of gene networks requires the development of computational methods for the automatic extraction of knowledge from tens of millions of scientific publications using machine learning and artificial intelligence approaches.
The aim of this work was to present the results of developing the cognitive system ANDSystem, which provides a full knowledge engineering cycle, from automatic analysis of scientific publications to gene network reconstruction.
Methods and Algorithms: Automatic knowledge extraction from natural language texts relied on a comprehensive framework integrating domain ontologies, machine learning methods for generating semantic-linguistic rules that enable rapid text analysis, and artificial intelligence methods, including graph neural networks and large language models.
Gene network reconstruction was carried out using original approaches based on an ontological description of the subject domain, including concepts and relationship types between biological entities.
Results: ANDSystem is a cognitive software and information system that considers multiple types of biological objects and extracts knowledge about diverse types of interactions from scientific texts [1–3]. The ANDSystem knowledge base, created through the automatic analysis of PubMed abstracts, PMC full-text articles, and factual databases, contains formalized knowledge on more than 150 million interactions between molecular genetic objects from different organisms, surpassing comparable systems worldwide in the number of supported object and relationship types.
In agrobiology, ANDSystem served as the basis for the development of the SmartCrop system for studying the molecular genetic mechanisms underlying genotype-phenotype relationships in agriculturally important rice and wheat crops [4]. In particular, SmartCrop has been applied to the identification of regulatory and signaling pathways involved in plant responses to specific growth conditions and to biotic and abiotic stresses, including high or low temperature, drought, salinity, soil contamination with metals, fertilizers, hormones, and elicitors. It has also been used for the prediction of candidate genes for genotyping, the search for markers for marker-assisted selection, and the identification of targets for the directed action of substances, including external factors.
In biomedicine, ANDSystem has been applied to the interpretation of omics data, the reconstruction of disease mechanisms, the analysis of drug-drug interaction mechanisms, the investigation of toxic effects of chemical compounds, the identification of drug targets, the discovery of novel biomarkers, and the detection of genetic risk factors for diseases. For example, ANDSystem has been used to study the molecular genetic mechanisms of pathogen-host interactions, reconstruct signaling pathways of cellular processes regulated by viral proteins, prioritize pharmacological targets for liver cancer therapy, investigate the regulatory role of SARS-CoV-2 non-structural proteins in metabolic disturbances in COVID-19 patients, and analyze blood plasma metabolomic screening data from patients with postoperative delirium to reconstruct the molecular genetic mechanisms of sphingolipid metabolism disorders.
Conclusion: ANDSystem is a powerful cognitive system for the reconstruction and analysis of gene networks with a unique knowledge base that surpasses similar known systems in the world. The system has been successfully applied to solve a wide range of problems in the field of agrobiology and biomedicine.
Funding: The bioinformatics analysis was supported by the budget project, FWNR-2022-0020.
Acknowledgements: The authors express their gratitude to Evgeniy V. Gaisler, Artem D. Rogachev, Nikita V. Basov, Andrey G. Pokrovsky for obtaining and analyzing the metabolomics data.

Список литературы/References

1. Ivanisenko V.A., Saik O.V., Ivanisenko N.V., Tiys E.S., Ivanisenko T.V., Demenkov P.S., Kolchanov N.A. ANDSystem: An Associative Network Discovery System for Automated Literature Mining in the Field of Biology. BMC Syst. Biol. 2015, 9 Suppl 2, S2.
2. Ivanisenko V.A., Demenkov P.S., Ivanisenko T.V., Mishchenko E.L., Saik O.V., A new version of the ANDSystem tool for automatic extraction of knowledge from scientific publications with expanded functionality for reconstruction of associative gene networks by considering tissue-specific gene expression. // 2019, BMC Bioinformatics., V.20(Suppl 1), H.34.
3. Ivanisenko TV, Demenkov PS, Ivanisenko VA. An Accurate and Efficient Approach to Knowledge Extraction from Scientific Publications Using Structured Ontology Models, Graph Neural Networks, and Large Language Models. Int J Mol Sci. 2024 Nov 3;25(21):11811. doi: 10.3390/ijms252111811.
4. Demenkov PS, Ivanisenko TV, Kleshchev MA, Antropova EA, Yatsyk IV, Volyanskaya AR, Adamovskaya AV, Maltseva AV, Venzel AS, Chao H, Chen M, Ivanisenko VA. SmartCrop: knowledge base of molecular genetic mechanisms of rice and wheat adaptation to stress factors. Vavilovskii Zhurnal Genet Selektsii. 2025 Dec;25(8):1221-1234. doi: 10.18699/vjgb-25-129