

РАЗРАБОТКА СИСТЕМЫ АВТОМАТИЧЕСКОГО АНАЛИЗА НАУЧНЫХ ПУБЛИКАЦИЙ ДЛЯ ОПРЕДЕЛЕНИЯ КЛЮЧЕВЫХ ПАРАМЕТРОВ ЛИОФИЛИЗАЦИИ МИКРООРГАНИЗМОВ

Павлишина И.Р. (ИТМО), Смирнов К.А. (ИТМО), Молчанов А.О. (ИТМО)
Научный руководитель – м.н.с., Зенкин А.М. (ИТМО)

Введение: Лиофилизация микроорганизмов является ключевым процессом для обеспечения стабильности и сохранения их биологической активности при длительном хранении. Оптимизация условий данного процесса имеет решающее значение для повышения эффективности биотехнологических приложений. Ряд исследований подчёркивает актуальность проблемы и демонстрирует возможности совершенствования технологии лиофилизации. Так, авторы работы [1] показали, что активация системы quorum sensing посредством N-ацил-гомосериновой лактона способствует улучшению длительного сохранения анаммокс микроорганизмов при вакуумной лиофилизации. Ранние исследования [2] продемонстрировали, что жизнеспособность лиофилизированных культур сохраняется даже после двадцатилетнего хранения. Кроме того, исследователи [3] выявили критическую роль водной активности и аморфного состояния для оптимального сохранения лактобактерий, а [4] предложили использование наностенных материалов для создания новых пробиотических продуктов. Дополнительные данные [5] свидетельствуют о возможности быстрой оптимизации формулировок для замораживания и сушки белков, в то время как [6] подтвердили перспективность современных подходов для длительного сохранения анаммокс бактерий. Эти результаты обуславливают необходимость разработки автоматизированной системы анализа научных публикаций для извлечения ключевых параметров лиофилизации, что и является целью данной работы.

Основная часть: Предложенная в исследовании система автоматического анализа научных публикаций включает три ключевых этапа: 1) *предварительная обработка текстовых данных*. На данном этапе осуществляется автоматическая фильтрация публикаций с выделением релевантной информации, содержащей ключевые запросы, что позволяет существенно сократить объём данных для дальнейшего анализа; 2) *извлечение информации*. С использованием больших языковых моделей [7–9] и алгоритмов машинного обучения происходит точное распознавание и извлечение параметров, характеризующих процесс лиофилизации. 3) *постобработка и структурирование данных*. Извлечённая информация преобразуется в удобный для анализа формат, что позволяет проводить сравнительный анализ и выявлять оптимальные условия для проведения лиофилизации.

Заключение: Представленный в работе метод автоматического анализа научных публикаций демонстрирует значительный потенциал для оптимизации процесса лиофилизации микроорганизмов. Применение данного подхода позволяет оперативно выявлять ключевые параметры, что способствует улучшению качества и долговременности хранения биоматериалов. Перспективным направлением дальнейших исследований является интеграция модулей для распознавания графической информации, что позволит расширить функциональные возможности системы и повысить точность извлекаемых данных.

Выполнено при поддержке Российского научного фонда (соглашение № 23-16-00224)

Список использованных источников:

1. Choi D., Sim B.-Y., Jung J. Activation of N-acyl-homoserine lactone-mediated quorum sensing system improves long-term preservation of anammox microorganisms by vacuum lyophilization // Chemosphere. 2022. Vol. 134743. DOI: 10.2139/ssrn.4035368.

2. Davis R. Viability and behavior of lyophilized cultures after storage for twenty-one years // Journal of Bacteriology. 1963. Vol. 85, № 2. P. 486–487. DOI: 10.1128/jb.85.2.486-487.1963.
3. Passot S., Cenard S., Douania I., Trelea I., Fonseca F. Critical water activity and amorphous state for optimal preservation of lyophilised lactic acid bacteria // Food Chemistry. 2012. Vol. 132. P. 1699-1705. DOI: 10.1016/J.FOODCHEM.2011.06.012.
4. Swe Z.M., Chumphon T., Panya M., Pangjit K., Promsai S. Evaluation of Nano-Wall Material for Production of Novel Lyophilized-Probiotic Product // Foods. 2022. Vol. 11. DOI: 10.3390/foods11193113.
5. Grant Y., Matejtschuk P., Dalby P. Rapid optimization of protein freeze-drying formulations using ultra scale-down and factorial design of experiment in microplates // Biotechnology and Bioengineering. 2009. Vol. 104. P. 957–964. DOI: 10.1002/bit.22448.
6. Rothrock M.J. Jr, Vanotti M.B., Szögi A.A., Gonzalez M., Fujii T. Long-term preservation of anammox bacteria // Applied Microbiology and Biotechnology. 2011. Vol. 92. P. 147-157
7. Gupta, T.; Zaki, M.; Krishnan, N. M. A.; Mausam. MatSciBERT: A Materials Domain Language Model for Text Mining and Information Extraction. NPJ. Comput. Mater. 2022, 8 (1), 102.
8. Polak, M. P.; Morgan, D. Extracting Accurate Materials Data from Research Papers with Conversational Language Models and Prompt Engineering. Nat. Commun. 2024, 15 (1), 1569.
9. Dagdelen, J.; Dunn, A.; Lee, S.; Walker, N.; Rosen, A. S.; Ceder, G.; Persson, K. A.; Jain, A. Structured Information Extraction from Scientific Text with Large Language Models. Nat. Commun. 2024, 15(1), 1418.

Автор _____ Павлишина И.Р.

Научный руководитель _____ Зенкин А.М.