

Тимофеев Г. А., Бердоносков В. Д.
G. A. Timofeev, V. D. Berdonosov

ОПТИЧЕСКОЕ РАСПОЗНАВАНИЕ СИМВОЛОВ В ИЗОЛИРОВАННОЙ ГИБРИДНОЙ ЭНЕРГЕТИЧЕСКОЙ СИСТЕМЕ: МЕТОДЫ АВТОМАТИЗАЦИИ СБОРА И ОБРАБОТКИ ДАННЫХ

OPTICAL SIMPLE DISCOVERING IN AN ISOLATED HYBRID ENERGY SYSTEM: METHODS OF AUTOMATIZATION OF DATA COLLECTION AND PROCESSING

Тимофеев Георгий Андреевич – аспирант Комсомольского-на-Амуре государственного университета (Россия, Комсомольск-на-Амуре).

Georgy A. Timofeev – Postgraduate Student, Komsomolsk-na-Amure State University (Russia, Komsomolsk-on-Amur).

Бердоносков Виктор Дмитриевич – кандидат технических наук, доцент кафедры «Прикладная математика» Комсомольского-на-Амуре государственного университета (Россия, Комсомольск-на-Амуре).

Victor D. Berdonosov – PhD in Engineering, Associate Professor, Applied Mathematics Department, Komsomolsk-na-Amure State University (Russia, Komsomolsk-on-Amur).

Аннотация. В статье рассматриваются методы автоматизации сбора и обработки данных об облачности для изолированных гибридных энергетических систем (ИГЭС) в поселениях Крайнего Севера, в том числе с использованием оптического распознавания символов (OCR). Основное внимание уделено разработке программных решений для сбора данных из метеорологических сервисов, таких как OpenWeather API, «маска облачности» от ДВО РАН и сервис Ventusky. Проведённый сравнительный анализ данных показал, что данные, полученные с сервиса Ventusky, обладают наибольшей корреляцией с реальными данными с установки, что делает его наиболее подходящим для использования в ИГЭС. Были разработаны новые методы, обеспечивающие более точное получение данных светопропускаемости и мониторинга системы. Созданный дашборд на основе библиотеки Dash обеспечивает визуализацию и мониторинг данных в реальном времени, способствуя оперативному принятию решений и дальнейшим исследованиям.

Summary. The paper considers methods of automation of cloud data collection and processing for isolated hybrid energy systems (IHES) in the Far North settlements, including the use of optical character recognition (OCR). The main attention is paid to the development of software solutions for data collection from meteorological services, such as OpenWeather API, cloud mask from FEB RAS and Ventusky service. Comparative data analysis showed that the data obtained from the Ventusky service has the highest correlation with real data from the installation, which makes it the most suitable for use in the IHES. New methods were developed to provide more accurate light transmission and system monitoring data. A dashboard based on the Dash library was created to provide real-time data visualization and monitoring, facilitating rapid decision-making and further research.

Ключевые слова: автоматизация сбора данных, оптическое распознавание символов, ИГЭС, поселения Крайнего Севера, OpenWeather API, маска облачности ДВО РАН, Ventusky, дашборд, гибридные энергетические системы.

Key words: data collection automation, optical character recognition, IHES, Far North settlements, OpenWeather API, cloud mask FEB RAS, Ventusky, dashboard, hybrid energy systems.

УДК 004.942

Введение. Точность данных об облачности критична для управления изолированными гибридными энергетическими системами (ИГЭС) в поселениях Крайнего Севера. В процессе создания автономной системы, которая функционирует на альтернативных источниках энергии и использует цифровой двойник для точного прогнозирования выработки энергии, важны точные данные об облачности [1]. Получение достоверной информации об облачности играет важную роль в

этой системе, т. к. влияет на эффективность прогнозирования и распределение ресурсов. Основная цель проекта – разработать и интегрировать методы автоматизации сбора и анализа данных об облачности (извлечение текстовых данных из изображений при помощи OCR-технологии) из различных метеорологических сервисов, таких как OpenWeather API [4], «маска облачности» от ДВО РАН [5] и Ventusky [6]. Эти методы помогут улучшить прогнозы солнечной энергии и управление ресурсами ИГЭС. Актуальность прогнозирования солнечной энергии подтверждается исследованиями, показывающими, что точный анализ данных об облачности повышает эффективность производства солнечной энергии [2; 3]. В ходе данной работы был проведён анализ существующих методов сбора данных об облачности и разработаны новые методы, обеспечивающие более точное получение данных светопропускаемости и мониторинга системы.

Обзор технологий и инструментов для интеграции данных в системы мониторинга.

Получение точных данных облачности представляет собой сложную задачу. В этом исследовании рассмотрены три источника данных об облачности: OpenWeather API [4], «маска облачности» от ДВО РАН [5] и Ventusky [6]. OpenWeather предоставляет доступ к широкому спектру метеорологических данных, включая информацию об облачности, температуре, влажности и скорости ветра. OpenWeather использует данные о погоде, в том числе информацию об облачности, от множества источников. В [4] сказано, что они используют открытые данные, а также с помощью свёрточных нейросетей и машинного обучения расширяют возможности прогнозирования, используют радиолокационные прогнозы, данные со спутников и метеостанций. В рамках исследования был реализован программный код, который позволяет автоматизировать процесс получения, сохранения и анализа этих данных. Программа использует API OpenWeather для получения исторических данных о погоде.

Второй источник данных представляет собой «маску облачности», разработанную Александром Андреевым и его коллегами из Дальневосточного центра ФГБУ «Научно-исследовательский центр космической гидрометеорологии “Планета”» и Вычислительного центра Дальневосточного отделения Российской академии наук. Исследование Андреева и Шамиловой посвящено методике детектирования облачности по данным геостационарного спутника Himawari-8 с применением нейронной сети свёрточного типа. Разработанная модель классификатора облачности была протестирована в различных сценариях, включая зимний и летний периоды года, в ночное и дневное время суток, а также в моменты смены дня и ночи. Результаты тестирования показали, что даже в сложных сценариях разработанный классификатор допускает минимальное количество ошибок по сравнению с применяемыми в мировой оперативной практике алгоритмами детектирования облачности. Обучение модели проводилось на данных с января 2016 года по июль 2019 года, включающих 302 мультиспектральных изображения территории Азиатско-Тихоокеанского региона [5]. После получения информации с источника была написана программа для разбора файлов формата «h5» и записи их в базу данных в «SQLite».

Третий источник данных для мониторинга облачности – приложение Ventusky [6], разработанное компанией InMeteo из Чехии совместно с Мареком Мойзиком и Мартином Прантлом. Ventusky представляет собой мощный инструмент для визуализации метеорологических показателей и изменения погодных условий в реальном времени по всему миру. Приложение предоставляет данные о ветре, давлении, осадках, облачности, температуре и грозах, отображая их на карте с использованием интуитивно понятных графических элементов. Основными поставщиками метеорологических данных для Ventusky являются Немецкая метеорологическая служба (DWD) и Национальное управление океанических и атмосферных исследований США (NOAA). Приложение собирает данные из различных источников, включая радары, спутники и наземные станции, чтобы предоставить точную и актуальную информацию о погоде. Получить данные с сервиса возможно при помощи технологии оптического распознавания символов (OCR) с использованием библиотеки «EasyOCR» [7] на Python. EasyOCR использует свёрточно-рекуррентную нейронную сеть CRNN для распознавания символов на изображениях. Алгоритм сначала сегментирует изображение, выделяя текстовые области, а затем классифицирует каждый символ в этих областях с помощью предварительно обученной модели. На основе данных со всех трёх источников был по-

строен объединённый график для сравнения информации об облачности и солнечной радиации (см. рис. 1).



Рис. 1. Сравнение облачного покрова и солнечной радиации с течением времени в декабре

Пунктирные линии разных цветов отображают данные об облачности: голубая линия — OpenWeather, серая линия — «маска облачности» ДВО РАН, оранжевая линия — Ventusky. Облачность измеряется в процентах, где 100 % означает полное облачное покрытие, а 0 — полное отсутствие облаков. Сплошная красная линия показывает данные о солнечной радиации, где высокие пики указывают на высокую интенсивность солнечной радиации, а низкие значения — на её отсутствие или низкую интенсивность. График показывает, что периоды высокой облачности соответствуют низким значениям солнечной радиации, что логично, т. к. облака препятствуют проникновению солнечных лучей. Периоды низкой облачности, напротив, характеризуются высокой солнечной радиацией. Как видно на графике, данные Ventusky имеют наибольшую корреляцию с реальными метеословиями.

Описание используемых методов и технологий для сбора и обработки данных. Процесс сбора данных начинается с формирования запросов к веб-приложению Ventusky при помощи базового URL с параметрами местоположения и времени. Для получения данных применяется автоматизация браузера с помощью библиотеки `pyautogui`. Программа открывает нужный URL, делает скриншоты области экрана с информацией об облачности и сохраняет текстовые данные, извлечённые из скриншота, в базу данных SQLite. Для распознавания текста из скриншотов используется библиотека `easyocr`.

Следующей задачей стала разработка интерфейса для визуализации и мониторинга данных. Необходимо было создать интерактивную панель управления (дашборд), которая позволит пользователям легко анализировать собранные данные об облачности и других погодных параметрах (см. рис. 2). Для создания дашборда была использована библиотека Dash, которая позволяет разрабатывать веб-приложения. Для программного решения получены соответствующие свидетельства о государственной регистрации [8–12]. В общем итоге полученный метод автоматизированного сбора метеорологических данных с использованием OCR (АСМ-OCR) позволил получать актуальную информацию об облачности и сохранять её в удобном формате для последующего использования.

На изображении показан веб-интерфейс дашборда для визуализации данных по солнечной радиации и облачности, созданный с использованием библиотеки Dash. Он содержит интерактивные элементы управления для анализа и мониторинга метеорологических данных, такие как кнопки для загрузки баз данных и выпадающие списки для выбора графиков. Дашборд позволяет пользователям анализировать данные, скачивать базы данных и способствует более точному прогнозированию и управлению ресурсами ИГЭС.

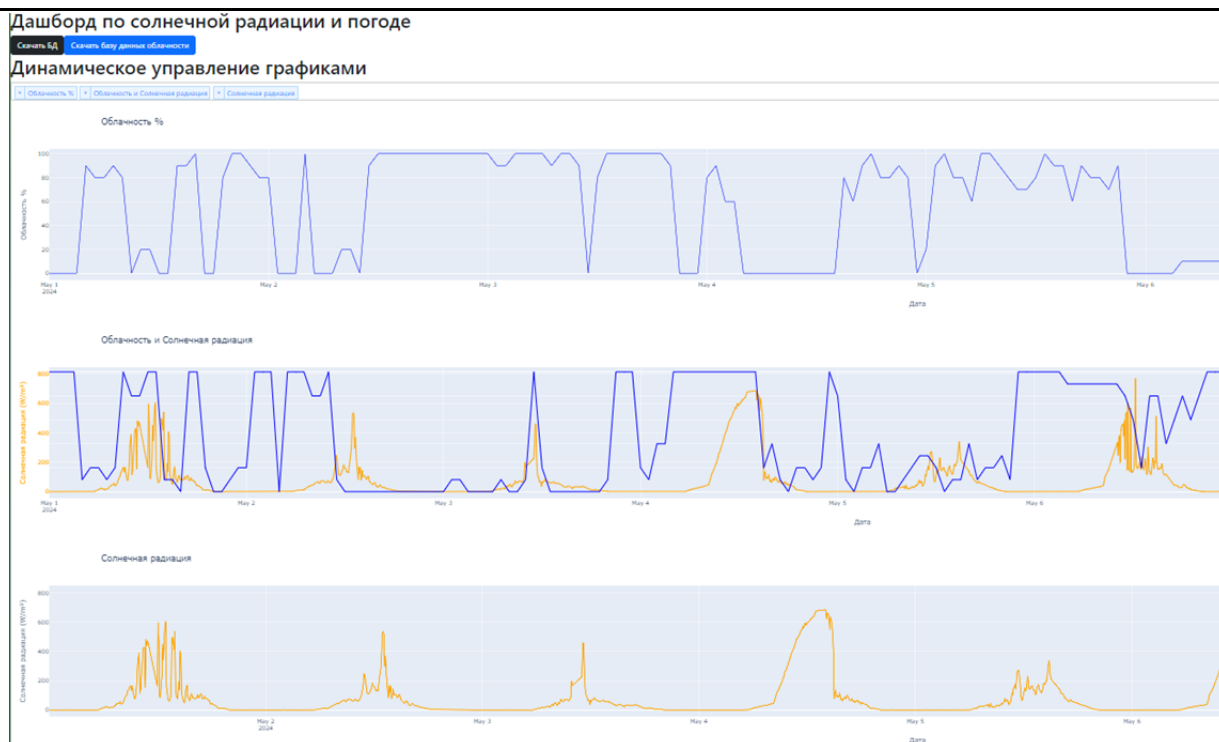


Рис. 2. Созданный интерфейс

Заключение. Сравнительный анализ показал, что данные Ventusky имеют наибольшую корреляцию с реальными метеорологическими условиями, что делает их наиболее подходящими для данного исследования. «Маска облачности» от ДВО РАН также продемонстрировала высокую точность, но имеет ограничения для мониторинга в реальном времени. Данные OpenWeather оказались менее точными для локальных условий. Разработанный метод ACM-OCR позволяет собирать, накапливать и визуализировать данные об облачности, обеспечивая удобство и гибкость в использовании. В итоге система мониторинга ИГЭС значительно улучшает точность и скорость обработки метеорологических данных, что важно для эффективного управления ресурсами в удалённых поселениях с суровыми климатическими условиями.

ЛИТЕРАТУРА

1. Тимофеев, Г. А. Поиск подходящей архитектуры для разработки цифрового двойника гибридных энергетических систем в изолированных от сетевых энергосистем средах с использованием ТРИЗ-эволюционного подхода / Г. А. Тимофеев // Вестник Новосибирского государственного университета. Серия: Информационные технологии. – 2022. – Т. 20. – № 4. – С. 76-99.
2. Информационные службы облачности для управления солнечной энергией в Западной Африке / Д. К. Дансо, С. Анкетин, А. Дьедхиу, Р. Адаму // Atmosphere. – 2020. – Том 11. – № 8. – URL: https://www.researchgate.net/publication/343645191_Cloudiness_Information_Services_for_Solar_Energy_Management_in_West_Africa (дата обращения: 23.06.2024). – Текст: электронный.
3. Cloudy Days and Solar Arrays // NESDIS, NOAA. – 2020. – URL: <https://www.nesdis.noaa.gov/news/cloudy-days-and-solar-arrays> (дата обращения: 23.06.2024). – Текст: электронный.
4. OpenWeatherMap // OpenWeather. – URL: <https://openweathermap.org> (дата обращения: 23.06.2024). – Текст: электронный.
5. Андреев, А. И. Детектирование облачности по данным КА Himawari-8 с применением свёрточной нейронной сети / А. И. Андреев, Ю. А. Шамилова // Исследование Земли из космоса. – 2021. – № 2. – С. 42-52.
6. Ventusky – ветер, дождь и температурные карты // Ventusky. – URL: <https://www.ventusky.com> (дата обращения: 23.06.2024). – Текст: электронный.

7. EasyOCR // Jaided AI. – URL: <https://github.com/JaidedAI/EasyOCR> (дата обращения: 23.06.2024). – Текст: электронный.
8. Свидетельство о государственной регистрации базы данных № 2024620134 Российская Федерация. Солнечно-Метеорологическая База Данных: № 2024620001: заявлено 05.01.2024: опубликовано 11.01.2024 / Тимофеев Г. А.
9. Свидетельство о государственной регистрации программы для ЭВМ № 2024610649 Российская Федерация. Получение данных с почты ИГЭС: № 2024610015: заявлено 05.01.2024: опубликовано 12.01.2024 / Тимофеев Г. А.
10. Свидетельство о государственной регистрации программы для ЭВМ № 2024610745 Российская Федерация. Программное решение для мониторинга экологических параметров в реальном времени для работы ИГЭС: № 2024610115: заявлено 05.01.2024: опубликовано 12.01.2024 / Тимофеев Г. А.
11. Свидетельство о государственной регистрации программы для ЭВМ № 2024610746 Российская Федерация. Программное получение показателей напряжения, тока, мощности со скриншотов для работы ИГЭС: № 2024610118: заявлено 05.01.2024: опубликовано 12.01.2024 / Тимофеев Г. А.
12. Свидетельство о государственной регистрации программы для ЭВМ № 2024612662 Российская Федерация. Программное получение метеорологической информации: № 2024610046: заявлено 05.01.2024: опубликовано 02.02.2024 / Тимофеев Г. А.