

**Ю. Ю. Красовская**

## ПАРАМЕТРИЗАЦИЯ МЕТАДААННЫХ АНГЛОЯЗЫЧНЫХ И РУССКОЯЗЫЧНЫХ ТРЕВЕЛ-БЛОГОВ

В современном цифровом пространстве тревел-блоги становятся одним из самых популярных и влиятельных форматов медиаконтента. Благодаря доступности цифровых технологий и онлайн-платформ тревел-блоггеры нередко выступают в роли конкурентов профессиональным журналистам и туристическим агентствам, оказывая влияние на поведение и выбор потребителей.

В отличие от медиатекстов традиционной журналистики, тревел-блоги характеризуются высокой степенью субъективности, гибкостью формата и мультимодальной природой, ориентированностью на массовую аудиторию (Т. Г. Добросклонская, 2008). В нашем исследовании под мультимодальностью понимаются способы восприятия информации одним из органов чувств (А. А. Кибрик, 2010). Мультимодальная природа тревел-блогов усиливает воздействие контента, формируя комплексное восприятие, ориентированное на массовую аудиторию. Автоматизированный анализ тревел-блогов вызывает трудности: тревел-контент отличается спонтанностью, эмоциональностью, насыщенностью элементами интернет-коммуникации (эмодзи, хэштеги, гиперссылки), а также наличием визуальных и аудиальных компонентов, значение которых зачастую сложно интерпретировать вне контекста.

Для автоматизированного анализа тревел-блогов необходимо выполнение ряда последовательных этапов, первым из которых является параметризация метаданных. Под параметризацией метаданных тревел-блогов понимается процесс выделения, структурирования и формализации вспомогательной информации о медиаконтенте, не относящейся непосредственно к основному тексту, но играющей значимую роль в его интерпретации и аналитической обработке. К метаданным тревел-блогов относятся имя или псевдоним автора, дата публикации, числовые показатели вовлечённости аудитории (лайки, просмотры, комментарии и т.п.), сведения о платформе размещения.

В рамках исследования нами разработан прототип программного обеспечения для автоматизированной параметризации мультимодального тревел-контента с платформ TikTok, YouTube, Instagram с использованием языка программирования Python. Для каждой из платформ разработан отдельный скрипт в зависимости от технических особенностей и ограничений платформы. Основной задачей разработанного программного обеспечения является сбор метаданных в автоматическом режиме.

Разработанный скрипт для платформы TikTok позволяет извлекать ключевые характеристики из короткого видео: URL-ссылку, имя автора, количество лайков, комментариев и сохранений. Наша программа оснащена графическим интерфейсом на базе библиотеки Tkinter, что обеспечивает

доступность её использования для пользователей без технической подготовки. После ввода ссылки на TikTok-видео скрипт инициирует браузер в фоновом режиме, осуществляет переход по указанному адресу и извлекает необходимые параметры с веб-страницы. Полученные данные автоматически сохраняются в Excel-таблице, с проверкой на наличие дубликатов, т.е. ранее обработанные видео не включаются повторно. Результатом работы скрипта становится структурированная база метаданных, пригодная для анализа пользовательской активности, уровня вовлечённости и тематических предпочтений в тревел-сегменте платформы.

Аналогичный скрипт создан нами для платформы YouTube с использованием библиотеки `yt_dlp` – форка `youtube-dl`. Скрипт предназначен для извлечения ключевых метаданных видеоконтента: названия видео, имени канала, количества просмотров, лайков, комментариев, даты загрузки. Эта программа также оснащена графическим интерфейсом, в котором пользователь вводит ссылку на видео, после чего скрипт с заданными параметрами (например, `noplaylist=True`, `extract_flat=True`, `quiet=True`) извлекает только метаданные без загрузки самого видеофайла. Полученные данные сохраняются в Excel-таблицу с проверкой на дублирование записей.

Для повышения эффективности работы скрипта предусмотрены дополнительные настройки, исключающие обработку плейлистов и подавляющие лишние предупреждения. Извлечение метаданных с платформы YouTube осложняется тем, что значительная часть информации загружается динамически посредством JavaScript и недоступна при стандартном HTML-парсинге. Дополнительные сложности связаны с мерами защиты YouTube от автоматизированных запросов: капчами, лимитами и блокировками.

Для извлечения метаданных с платформы Instagram нами создан аналогичный скрипт с использованием библиотеки `instaloader`, которая позволяет имитировать действия пользователя и позволяет осуществлять обход ограничений платформы. При вводе пользователем ссылки на пост программа извлекает `shortcode`, по которому с помощью метода `Post.from_shortcode` формируется объект публикации. Из него автоматически извлекаются ключевые параметры: имя автора, количество лайков, дата публикации, а также тип контента (фото или видео). Полученные данные сохраняются в Excel-таблицу и могут использоваться для дальнейшего анализа.

Извлечение метаданных с платформы Instagram сопряжено с рядом ограничений, связанных с политикой конфиденциальности и защитой данных от автоматизированного доступа. Доступ к большинству метаданных возможен только при наличии авторизации. Платформа активно использует механизмы защиты от парсинга, включая CAPTCHA, ограничения по IP-адресам и требование верификации через SMS или электронную почту, что существенно осложняет автоматизацию процесса. Дополнительные сложности связаны с ограниченным доступом к официальному API Instagram, который требует авторизации через OAuth, установленных лимитов на количество запросов и регистрации приложений.

Таким образом параметризация метаданных позволила систематизировать большие массивы неструктурированной информации и создать единую модель описания как англоязычных, так и русскоязычных тревел-блогов. Это, в свою очередь, делает возможным в дальнейшем сопоставительный анализ, включая оценку популярности публикаций, динамики интереса к темам и регионам, активности авторов и особенностей восприятия путешествий в различных языковых и культурных средах.