АНАЛИТИКА БОЛЬШИХ ДАННЫХ ДЛЯ ПОИСКОВОЙ ОПТИМИЗАЦИИ

Илиусыз Алмат Санатулы

магистрант, Евразийский национальный университет имени Л.Н. Гумилева, Казахстан, г. Нур-Султан

Тусупов Джамалбек Алиаскарович

научный руководитель, д-р физ.-мат. наук, профессор, Евразийский национальный университет имени Л.Н. Гумилева, Казахстан, г. Нур-Султан

В эпоху больших данных поисковая оптимизация имеет дело с инкапсуляцией наборов данных, связанных с производительностью веб-сайта с точки зрения архитектуры, курирования контента и поведения пользователей, с целью преобразования их в действенные идеи и улучшения видимости и доступности в Интернете. В этом отношении аналитика больших данных расширяет возможности для разработки новых методологических рамок, состоящих из достоверной, надежной и последовательной аналитики, которая практически полезна для разработки хорошо обоснованных стратегий оптимизации органического трафика. В данной работе реализована новая методология для увеличения числа посещений в органических поисковых системах, основанная на влиянии множества SEO-факторов. Для достижения этой цели авторы исследовали 171 сайт культурного наследия и полученные на нем аналитические данные об их работе и пользовательском опыте внутри них. Огромное количество веб-коллекций включено и представлено организациями культурного наследия через их веб-сайты. Впоследствии пользователи взаимодействуют с этими коллекциями, производя поведенческую аналитику в различных типах данных, поступающих с нескольких устройств, с высокой скоростью и в больших объемах. Тем не менее, предыдущие исследования показывают, что эти массивные культурные коллекции трудно просматривать, выражая низкую видимость и доступность в эпоху семантической сети. На этом фоне в данной статье предлагается вычислительная разработка стратегии поисковой оптимизации (SEO), которая использует сгенерированную аналитику больших культурных данных и улучшает видимость сайтов культурного наследия. Далее статистические результаты исследования интегрируются в прогностическую модель, состоящую из двух этапов. Во-первых, процесс нечеткого когнитивного отображения генерируется как агрегированная описательная модель макроуровня. Во-вторых, за этим следует модель на микроуровне, основанная на данных агентов. Целью модели является прогнозирование наиболее эффективных комбинаций факторов, обеспечивающих повышенную видимость и органичный трафик на сайтах организаций культурного наследия. С этой целью исследование способствует расширению знаний исследователей и практиков в секторе большой культурной аналитики с целью реализации потенциальных стратегий повышения видимости и доступности культурных коллекций в Интернете.

За последние 10 лет аналитику больших данных стали называть “нефтью” для оптимизации цифровой экосистемы, а впоследствии и сферы Всемирной паутины. Необузданные большие данные с огромным объемом и скоростью генерируются относительно взаимодействия пользователей с поисковыми системами и веб-сайтами и того, как они соответственно реагируют на результаты поиска и контент, который они получают. Поисковые системы, результаты поиска и веб-сайты выражают причинно-следственную связь в соответствии с их основной целью существования.

Эта цель связана с предоставлением наибольшего объема информации, в лучшие сроки, с наиболее точными результатами, в соответствии с поисковыми запросами пользователей.

В контексте эпохи больших данных поисковая оптимизация (SEO) играет решающую роль в потенциальном распространении персонализированного контента, отражающего качество. Это качество связано с курированием контента и надлежащим юзабилити в веб-системах для удовлетворения информационных потребностей пользователей. Основная цель SEO связана с предоставлением стратегических шагов веб-разработчикам и создателям контента для оптимизации веб-сайтов для более высокого рейтинга в результатах поиска и, следовательно, большего органического трафика, поступающего из поисковых систем.

Тем не менее, несмотря на большой объем наборов данных и аналитики, которые производятся и связаны с производительностью веб-сайтов и поведением пользователей внутри них, область исследований по использованию аналитики больших данных для стратегических схем SEO все еще находится в зачаточном состоянии. С этим пробелом в исследованиях связано несколько причин, таких как симптомы перегрузки данных и, следовательно, трудности предварительной обработки, анализа, визуализации и интерпретации результатов. Впоследствии еще один недостаток связан с уменьшением управленческих возможностей для понимания взаимосвязей между метриками семантической веб-аналитики и тем, как они согласуются с целями SEO. Еще одним трудным моментом является отсутствие методологических механизмов, которые четко формулируют валидность, надежность и согласованность в отношении переменных, которые принимаются во внимание с целью оптимизации видимости веб-сайтов. На этом фоне в данной статье представлен новый методологический подход к использованию аналитики больших данных, связанной с производительностью веб-сайта и тем, как они способствуют достижению цели SEO, которая заключается в увеличении процента органического трафика поисковых систем. Мы реализуем эту методологию в области культурного наследия.

Институты культурного наследия (ЧИС) и их содержание как в физическом, так и в цифровом мире представляют собой социальное стремление сохранить и связать наследие прежних поколений в современном мире, отражая родовую идентичность каждого общества. Культурные учреждения, такие как музеи, галереи, библиотеки и архивы, поддерживают демократизацию культурного наследия. Они формируют и укрепляют культурный и образовательный фон народа, в то время как их миссия состоит в том, чтобы объединить разрозненные фрагменты истории в зависимости от общества, к которому они принадлежат и которому служат. Информационно-коммуникационные технологии и Интернет расширяют возможности учреждений культуры привлекать больше заинтересованных сторон как в физической, так и в онлайн-среде. Этот факт усиливает их возможности для предоставления культурных знаний заинтересованным сторонам, глобальным, национальным и местным сообществам, обучая их довольно привлекательным способом через веб-сайты. Это предполагает переход от традиционных подходов keep and protect к опыту и участию в эпоху цифровых гуманитарных наук.

Согласно Уилсону, в каждом учреждении культурного наследия (УКН) первой основной целью является управление содержащимся в них культурным материалом. Вторая цель-сделать этот материал доступным для всех, кто хочет его увидеть. Действительно, даже начиная с предыдущих исследований на стадии зарождения Интернета и вплоть до недавних подходов, ЧИС использовали возможности веб–присутствия через веб-сайты для расширения видимости и доступности контента своего культурного наследия. Веб-сайты под отцовством УКН представляют собой жизненно важный онлайн-канал, который повышает осведомленность и расширяет возможности для посещения. В то же время, это дает преимущество представления культурной информации, которая способна намного превосходить ту, которая доступна в физическом месте. Впоследствии исследования Ворбий и Марти показали, что пользовательский опыт на разных веб-сайтах приводит к лучшей подготовке посетителей, а также к улучшению опыта после посещения с точки зрения значения и интерпретации, которые они приписывают культурным коллекциям.

Однако использование веб-сайтов для УКН с целью оптимизации видимости их артефактов в Сети-это не простой и легкий путь. Веб-сайты УКН состоят из огромного количества наборов данных, связанных с архитектурными аспектами, такими как большое количество веб-страниц, которые они содержат, сотни изображений, тысячи внутренних и внешних взаимосвязей и т.д. В этой связи предыдущие усилия отмечали, что тенденция роста оцифровки артефактов не является достаточной или организованной для точного поиска

культурной информации в крупномасштабных веб-коллекциях. Более конкретно, это ставит перед посетителями задачу столкнуться с (а) симптомами перегрузки предоставленной информации о культурном наследии и/или (б) отсутствием культурной информации из-за низкого уровня доступности и видимости.

Эти две проблемы, упомянутые выше, отражают нехватку эффективного SEO-подхода, который можно было бы количественно измерить для повышения точности поиска информации с веб-сайтов УКН. Это заявление также было поддержано недавним докладом Крстича и Масликовича. Они показали, что только треть обследованных ими учреждений культуры подтвердили использование аналитических платформ для оценки и оптимизации пользовательского опыта при поиске контента и навигации по цифровым коллекциям. Этот вид оценки, предназначенный для усиления контента, который страдает от низкого уровня доступности и видимости, становится еще более сложным, поскольку веб-сайты УКН характеризуются огромными размерами своего контента.

Список литературы:

1. Ди Франко, П. Д. Г.; Мэтьюз, Дж. Л.; Мэтлок, Т. Обрамление прошлого: как виртуальный опыт влияет на телесное описание артефактов. Herit. 2016, 17, 179-187. [CrossRef]

2. Крстич Н.; Масликович Д. Болевые точки культурных учреждений в поиске видимости: на примере Сербии. Либр. Hi Tech News 2018, 37, 496-512.