

Научная статья

УДК 004.92

doi: 10.34822/1999-7604-2022-1-29-37

РЕШЕНИЕ ЗАДАЧ ИНТЕРНЕТ-МАРКЕТИНГА СРЕДСТВАМИ PYTHON

Имран Мусаевич Маггеррамов^{1✉}, Гурру Имранович Акперов²

^{1,2}Южный университет (ИУБуП), Ростов-на-Дону, Россия

¹imr.magerramov@yandex.ru[✉], <http://orcid.org/0000-0003-3179-5742>

²pr@iubip.ru, <http://orcid.org/0000-0002-9057-0001>

Аннотация. Рассмотрены вопросы подготовки и визуализации данных для целей интернет-маркетинга с использованием технологий Python (предварительная обработка данных от поставщиков и формирование итоговых массивов для отчетов) и PowerBI (визуализация данных для целей дальнейшей интерпретации и принятия решений). Представлено решение проблемы анализа эффективности продвижения страниц сайтов в поиске Google. Методическим аппаратом для реализуемых алгоритмов выступает набор ключевых показателей эффективности (показы, клики, CTR, средневзвешенная позиция, объем и процент конверсий, объем прибыли и пр.), а также аппарат нечеткой логики (выделение приоритетных к продвижению страниц по заранее определенным критериям).

Ключевые слова: нечеткая логика, объектно-ориентированное программирование, визуализация и интерпретация данных, обработка данных, хранение данных, BI-системы

Для цитирования: Маггеррамов И. М., Акперов Г. И. Решение задач интернет-маркетинга средствами Python // Вестник кибернетики. 2022. № 1 (45). С. 29–37. DOI 10.34822/1999-7604-2022-1-29-37.

Original article

SOLVING INTERNET MARKETING PROBLEMS VIA PYTHON

Imran M. Magerramov^{1✉}, Gurru I. Akperov²

^{1,2}Southern University (IMBL), Rostov-on-Don, Russia

¹imr.magerramov@yandex.ru[✉], <http://orcid.org/0000-0003-3179-5742>

²pr@iubip.ru, <http://orcid.org/0000-0002-9057-0001>

Abstract. The article discusses the issues of data pre-processing and visualization for internet marketing using such technologies as Python (front-end processing of data received from suppliers and formation of datasets for reports) and PowerBI (visualization of data for further interpretation and decision-making purposes). The problem of analysis of efficiency of website pages promotion in Google search was solved. A set of key performance indicators (impressions, clicks, CTR, an average weighted position, the amount and rate of conversions, the amount of profit etc.) and the fuzzy logic paradigm (detection of pages preferred for promotion according to the pre-determined criteria) were used as methods for the implemented algorithms.

Keywords: fuzzy logic, object oriented programming, data visualization and interpretation, data processing, data storage, BI-systems

For citation: Magerramov I. M., Akperov G. I. Solving Internet Marketing Problems via Python // Proceedings in Cybernetics. 2022. No. 1 (45). P. 29–37. DOI 10.34822/1999-7604-2022-1-29-37.

ВВЕДЕНИЕ

Цель оценки эффективности страниц сайта в поиске – выделение дальнейшего фронта работ по оптимизации ресурса по ряду выведенных критериев. Для оценки небольшого количества ресурсов (до 10–15) разной направленности всю необходимую информацию для анализа можно получить в основных системах веб-аналитики (Яндекс, Метрика и Google Search Console). Однако при значительно большем количестве ресурсов, а также пересечении их тематики у команд оптимизации возникает потребность в консолидации данных в едином инструменте для быстрого отслеживания динамики изменений ключевых показателей и принятия решений. В данной работе представлен механизм создания такого инструмента средствами Python и PowerBI.

МАТЕРИАЛЫ И МЕТОДЫ

В предыдущих работах описан механизм оценки альтернатив с помощью методики нечеткого бенчмаркинга [1–3]. Этот методический аппарат планируется использовать и для автоматизации финального этапа оценки – выделения страниц, которые необходимо продвигать в первую очередь.

По архитектурным соображениям для визуализации данных выбрана среда Microsoft PowerBI, имеющая широкие интеграционные возможности, а также функционал датасетов (поточных наборов данных для их прямой передачи в отчеты как с возможностью исторического анализа, так и без нее) [4]. Рассматриваемая в качестве альтернативы система Google Data Studio подходит только для визуализации данных по принципу «один ресурс – одна витрина данных», что неприемлемо для решения поставленной в данной работе задачи [5].

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Формулировка задачи и начальных условий. В качестве примера рассмотрим следующую задачу: имеется сеть из 62 доменов, необходимо построить систему консолидированной аналитики изменений по этим ресурсам в динамике по следующим показателям (в разрезе по каждой странице ресурса):

- поисковые (показы, клики, CTR, средняя взвешенная позиция);
- величина понесенных расходов на страницу в периоде;
- результирующие показатели (объем регистраций, процент конверсий, объем и сумма депозитов, сумма прибыли, рентабельность);
- географические показатели (регион, страна);
- непосредственно ресурс, к которому относится страница.

На выходе необходимо получить отчет по страницам, выделить приоритетные и собрать для них запросы по заданным критериям (показы более 10; позиция от 8 до 30; запрос не содержит прямого упоминания бренда).

Доступные срезы:

- М/М: расходы и положительный эффект в этом месяце;
- All Time: расходы в этом месяце, положительный эффект с начального периода (август 2018 г.).

Источники данных:

- CRM-система (результирующие показатели по страницам и расходы на их продвижение представлены облачными таблицами в Google Spreadsheets);
- статистика по сайтам в Google Search Console.

Подготовка инфраструктуры. На стороне PowerBI создадим потоковые наборы данных со структурой для указанных срезов (рис. 1) согласно следующей логике: по каждой странице система будет ожидать указанный набор полей. Прием данных осуществляется с помощью POST-запроса на URL [2] обработки пополнения набора данных (URL отображается в настройках набора данных и при его создании). Создадим такие наборы для обоих типов основных отчетов – All Time и М/М.

Далее реализуем логику сбора, обработки и отправки данных от информационных объектов региона [6] в наборы для основных отчетов.

В первую очередь необходимо определить список ресурсов. Реализуем абстрактный класс GetSites для работы с этим списком. Рассмотрим его логику на рис. 2.

```
[
{
"Сайт" : "AAAAA555555",
"Страница" : "AAAAA555555",
"Клики" : 98.6,
"Показы" : 98.6,
"CTR" : 98.6,
"Позиция" : 98.6,
"Регистрации" : 98.6,
"CR, %" : 98.6,
"Суммы депозитов" : 98.6,
"FTD" : 98.6,
"% FTD" : 98.6,
"Страна" : "AAAAA555555",
"Регион" : "AAAAA555555",
"Расходы" : 98.6,
"ROI, %" : 98.6,
"PnL" : 98.6,
"PnL на регистрацию" : 98.6,
"Бренд" : "AAAAA555555",
"Месяц" : "2021-11-19T22:36:08.258Z",
"Тематика" : "AAAAA555555",
"Тип страницы" : "AAAAA555555"
}
]
```

Рис. 1. Структура данных набора для основного отчета

Примечание: скриншот авторов.

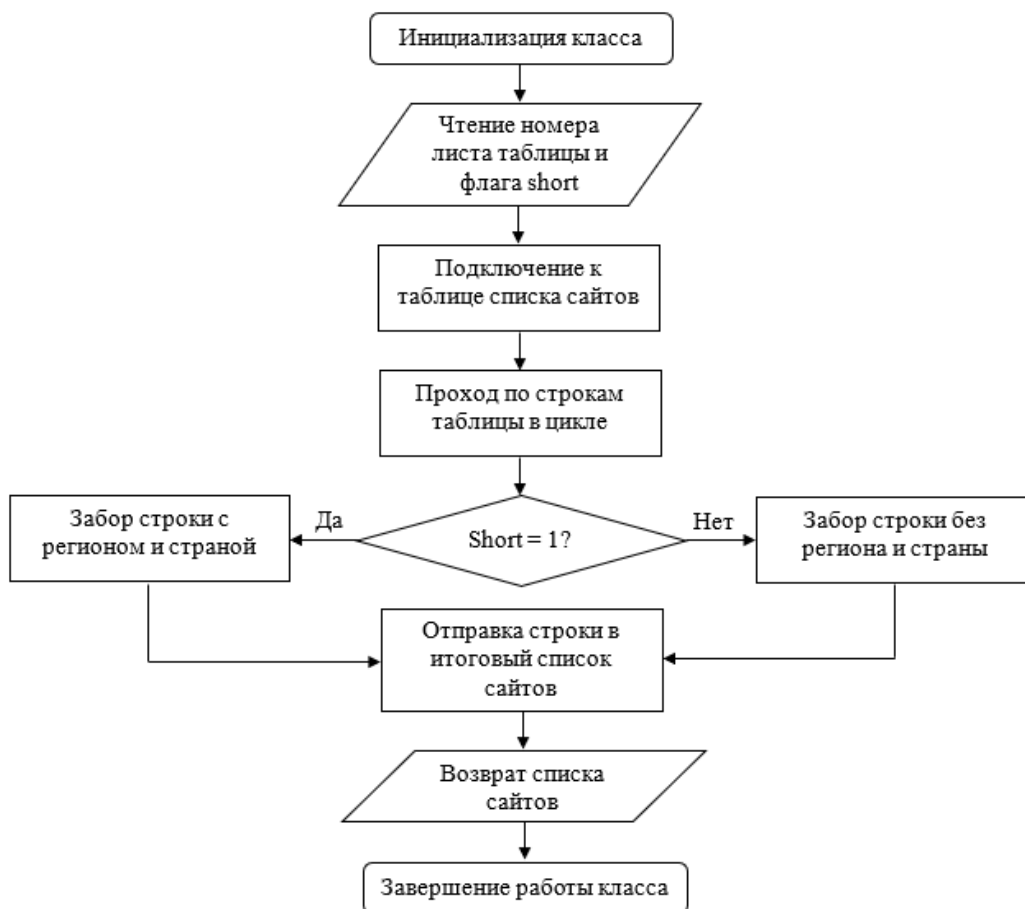


Рис. 2. Класс для работы со списком ресурсов

Примечание: составлено авторами.

В классе, помимо конструктора, содержится публичная функция `fill()`, в логике которой описана работа с Excel-файлом списка сайтов. Класс принимает название листа Excel-книжки, а также маркера, указывающего на необходимость забора (и агрегирования) [7–8] региональной привязки сайта или же ее отсутствие. В функции `fill()` реализован забор нового

и старого доменного имени сайта (для ситуации переезда и необходимости дальнейшего суммирования статистики с двух доменов). На выходе функция `fill()` возвращает список сайтов для сбора статистики. Далее реализуем класс для работы с облачной таблицей данных CRM (рис. 3).

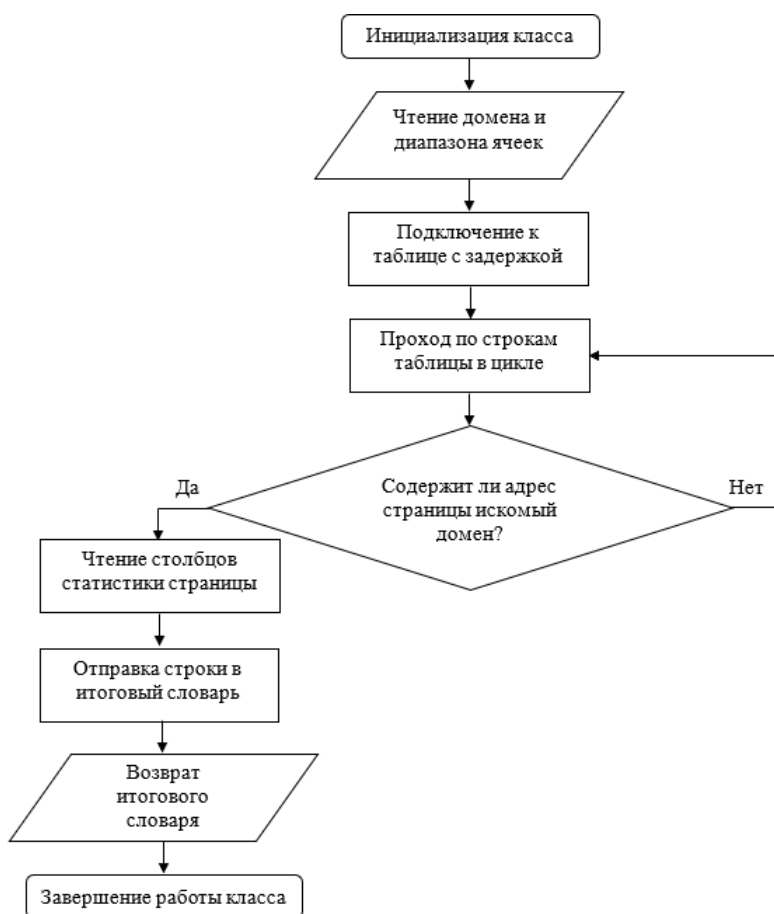


Рис. 3. Класс для работы со статистикой CRM

Примечание: составлено авторами.

Принимая на входе доменное имя и диапазон ячеек таблицы, класс подключается к ней по API Google Spreadsheets и заполняет итоговый словарь данными, используя в качестве ключа страницу. Подключение к таблице выполняется при каждом вызове функции `fill()`, в связи с чем для предотвращения блокировки в классе реализована приватная функция `__setDelay()`, устанавливающая между запросами задержку в 8 сек.

Аналогичным образом реализуем класс для работы с таблицей расходов. Он будет содержать приватную `__setDelay()` и публичную `collect()` функцию, внутри которой будет происходить заполнение итогового словаря `upload` расходами по каждой странице. Конструктор класса принимает ресурс, имя листа в таблице и диапазон ячеек для обхода.

Далее необходимо реализовать логику для работы с API Google Search Console (рис. 4).

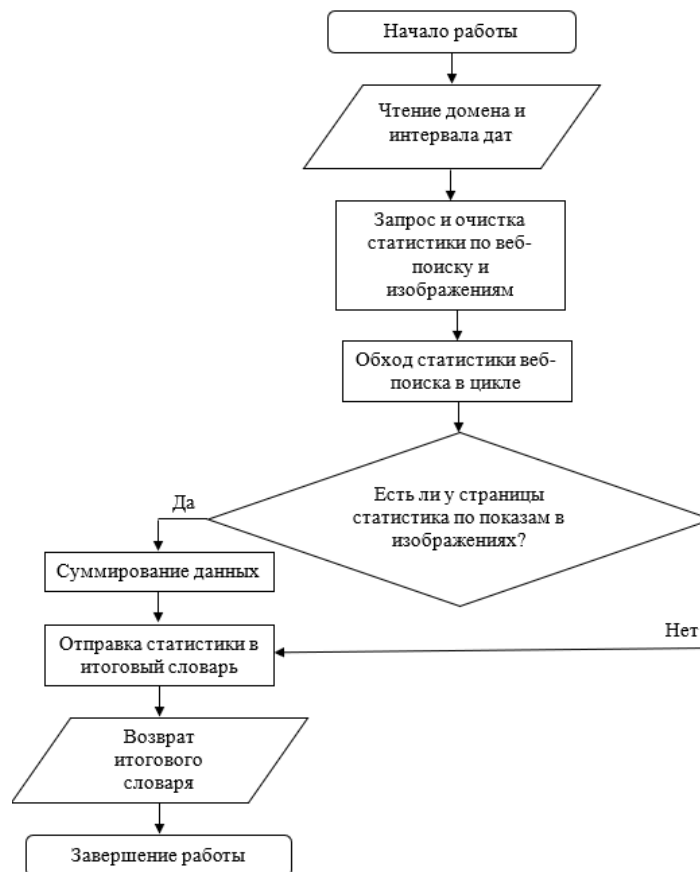


Рис. 4. Алгоритм работы с данными статистики Google по страницам
Примечание: составлено авторами.

Приведенная логика забирает из Google Search Console постраничную статистику по одному сайту (свойство `property_url` в объекте `argv`), очищает ее от лишних ключей и собирает в итоговый словарь с итерированием по каждой странице [9]. Поскольку в статистике Google ведется учет показов на нескольких площадках (веб-поиск, картинки, новости и пр.), реализовано суммирование данных по трафику с последующим расчетом общего CTR и средневзвешенной позиции (функция `SumSEOParams()`).

Функция `filterSpecs()` позволяет отобрать из общей статистики страницы, удовлетворяющие ряду фильтров (для выделения неучитенных страниц в разрезе расходов и данных по CRM).

Далее реализуем абстрактный класс для соединения данных из всех источников и их отправки в отчет PowerBI (рис. 5).

Согласно приведенной логике ее инициализация выполняется при вызове функций `loadSpecs()` или `fetchStats()`. Конструктор класса принимает на входе листы и диапазоны ячеек статистики по CRM и расходам, начальную дату периода и URL потокового набора данных для отправки итогового массива в отчет. При этом реализовано суммирование статистики через внешнюю функцию `checkMark()`, а также созданы функции как для работы с основной статистикой (по данным CRM), так и с дополнительной (фильтрации по объему кликов, вхождению в таблицы по CRM и расходам или их отсутствию). На выходе функцией `__fillResult()` заполняется итоговый словарь данных, который преобразовывается в JSON и отправляется в отчет внешней функцией `handleRequestToBI()`.

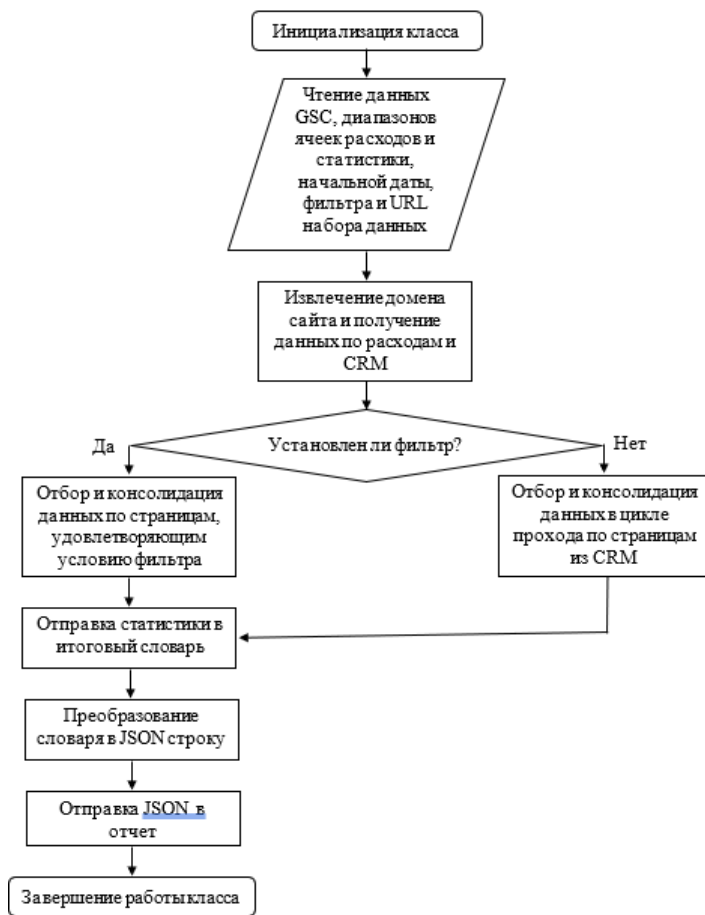


Рис. 5. Алгоритм консолидации и отправки данных в отчет
 Примечание: составлено авторами.

Далее обернем всю имеющуюся логику в оконное приложение, используя возможности библиотеки Tkinter. На рис. 6 пред-

ставлен вид интерфейса инструмента для загрузки.

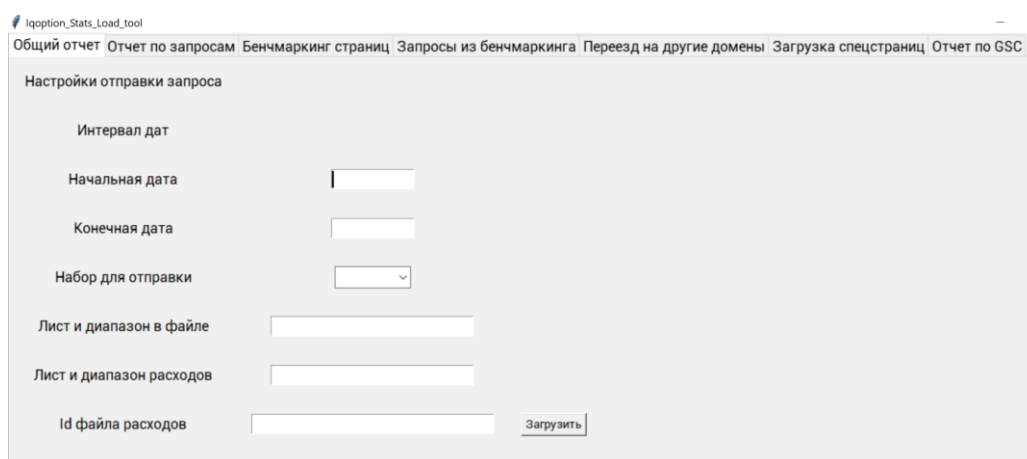


Рис. 6. Вид приложения
 Примечание: скриншот авторов.

В рамках данной работы процедура загрузки будет выполняться с помощью вкладки «Общий отчет» и «Загрузка спецстраниц». Остальные вкладки ориентированы на

загрузки.

использование мягких моделей [10–11] и будут описаны в последующих работах.

Пользователю необходимо задать начальную и конечную дату интервала, выбрать набор для отправки (All Time или M/M), определить диапазон ячеек в таблицах со ста-

тистикой по CRM и расходам и задать Id-таблицы расходов (если они хранятся не в той же таблице, что и данные по CRM).

Вид вкладки спецстраниц представен на рис. 7.

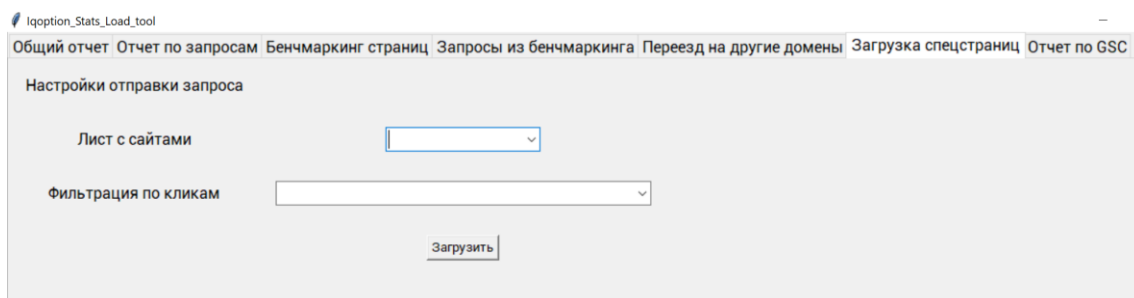


Рис. 7. Загрузка спецстраниц
Примечание: скриншот авторов.

Спецстраницы – это страницы, прошедшие через функции фильтрации, описанные в классе ExportStats. Они загружаются в общий отчет в отдельном потоке, для этого пользователю необходимо выбрать интересующий лист из списка сайтов, а также нужный тип фильтрации в соответствующих выпадающих списках.

Построение отчетов. Запустим приложение и выполним загрузку данных в наборы. После этого создадим отчеты по месячной статистике. Вид общей сводки по страницам приведен на рис. 8.

Страница		% FTD		Страна		Бренд		6522				
Search		0,00	100,00	Все		Все		Всего строк				
CR %		ROI %		Регистрации		Расходы		Месяц				
0,00		-10 344,19		0,00		0,00		Все				
112 800,00		68 625,91		139 185,00		944,73						
Месяц	Страница	Показы	Клики	CTR	Позиция	Регистрации	FTD	Суммы депозитов	RnL	RnL на регистрацию	Расходы	ROI, %
01.01.21 12:00:00		1 029 278,00	45 054,00	4,38	4,53	0,00	0,00	0,00	0,00	0,00	0,00	0,00
02.01.21 12:00:00		1 024 658,00	40 190,00	3,92	5,25	0,00	0,00	0,00	0,00	0,00	0,00	0,00
03.01.21 12:00:00		1 251 912,00	35 611,00	2,84	4,75	0,00	0,00	0,00	0,00	0,00	0,00	0,00
04.01.21 12:00:00		1 203 980,00	24 985,00	2,08	5,68	0,00	0,00	0,00	0,00	0,00	0,00	0,00
05.01.21 12:00:00		1 080 016,00	21 847,00	2,02	5,62	0,00	0,00	0,00	0,00	0,00	0,00	0,00
06.01.21 12:00:00		998 140,00	21 294,00	2,13	5,25	0,00	0,00	0,00	0,00	0,00	0,00	0,00
10.01.20 12:00:00		844 755,00	13 453,00	1,59	6,72	0,00	0,00	0,00	0,00	0,00	0,00	0,00
12.01.20 12:00:00		719 036,00	12 941,00	1,80	5,88	0,00	0,00	0,00	0,00	0,00	0,00	0,00
03.01.21 12:00:00		117 301,00	12 237,00	10,43	3,88	0,00	0,00	0,00	0,00	0,00	0,00	0,00
06.01.21 12:00:00		90 994,00	11 764,00	12,93	4,22	0,00	0,00	0,00	0,00	0,00	0,00	0,00
04.01.21 12:00:00		97 881,00	10 856,00	11,09	3,95	0,00	0,00	0,00	0,00	0,00	0,00	0,00
01.01.21 12:00:00		114 106,00	10 832,00	9,49	3,28	0,00	0,00	0,00	0,00	0,00	0,00	0,00
11.01.20 12:00:00	883 397,00	10 437,00	1,18	7,85	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
02.01.21 12:00:00	103 200,00	10 329,00	10,01	4,13	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
11.01.20 12:00:00	77 630,00	9 864,00	12,71	3,53	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
Всего		68 003 152,00	1 489 022,00			472 732,00	5 413,00	1 208 067,38	586 471,74		68 446,44	

Рис. 8. Сводка по страницам в Microsoft PowerBI
Примечание: скриншот авторов.

Манипулируя фильтрами, пользователь отчета может выделять страницы, удовле-

творяющие текущим критериям. Для целей дальнейшей автоматизации отчетности были

созданы копии страниц сводки с настроенными фильтрами. Страницы из этих фильтров используются для наполнения сводки по запросам в другом отчете, логика которого будет описана в последующих работах.

Обладая сводкой, пользователи отчетов могут строить любые срезы для нужд заинтересованных лиц.

ЗАКЛЮЧЕНИЕ

По итогу работы со сводкой были получены в целом положительные отзывы от команды оптимизации. На будущее был согласован следующий фронт изысканий:

- оптимизация логики загрузки на получение данных по всем страницам из Google Search Console;
- реализация нечеткой классификации страниц для дополнения информации, получаемой из фильтров по сводке [12];
- реализация дополнительных страниц в отчете для удобства экспорта данных;
- реализация листов с прогнозными графиками ключевых показателей;

- предварительное сохранение данных в СУБД и перевод отчетов на новую архитектуру (поскольку набор данных ограничен объемом в 200 000 строк, которых со временем может оказаться недостаточно) [2].

Представлены результаты визуализации данных для целей интернет-маркетинга на основе анализа данных ключевых показателей эффективности с применением аппарата нечеткой логики.

В работе рассмотрено создание консолидированной витрины данных по показателям списка сайтов. Предложена логика загрузки на Python и создание самих витрин в среде PowerBI. В результате реализации описанных изысканий получен реальный опыт решения бизнес-задач консолидации плохо структурированных данных и настройки их аналитики.

В дальнейшем планируется описание реализации подобных отчетов для последующих частей задачи – классификации страниц [1, с. 125] и сбора запросов по ним.

Список источников

1. Акперов Г. И., Магеррамов И. М., Бочаров А. А., Гребенюк Е. В. Оценка комплексной безопасности системы управления вузом на основе мягких моделей // *Соврем. информ. технологии и ИТ-образование*. 2021. Т. 17, № 1. С. 120–128.
2. Akperov I. G., Akperov G. I., Alekseichik T. V. et al. *Soft Models of Management in Terms of Digital Transformation*. Rostov-on-Don : PEI HE SU (IMBL), 2019. 188 p.
3. Магеррамов И. М. Моделирование принятия решений в условиях нечеткости требований // *Интеллектуальные ресурсы – региональному развитию*. 2021. Т. 7, № 1. С. 157–163.
4. Документация по PowerBI. URL: <https://docs.microsoft.com/ru-ru/power-bi/> (дата обращения: 20.01.2022).
5. Студия данных Google. Подключение и визуализация данных в Студии данных. URL: <https://developers.google.com/datastudio?hl=ru> (дата обращения: 20.11.2021).
6. Kramarov S. O., Khramov V. V., Belyaev A. V., Grebenyuk E. V. The Formalization of the Characteristics of Automatic Workplace (AWP) with the Use of Linguistic Variables (LV) // *IOP Conference Series: Materials Science and Engineering*, September 11–13, 2020. Rostov-on-Don, 2021. Vol. 1029. P. 012082. DOI 10.1088/1757-899X/1029/1/012082.
7. Чернышев Ю. О., Храмов В. В. Особенности агрегирования качественных признаков опорных ориентиров в системах технического зрения // *Известия ТРТУ*. 2001. № 3 (21). С. 55.

References

1. Akperov G. I., Magerramov I. M., Bocharov A. A., Grebenyuk E. V. Assessment of the Integrated Safety of the University Management System Based on Soft Models // *Modern Information Technologies and IT Education*. 2021. Vol. 17, No. 1. P. 120–128. (In Russian).
2. Akperov I. G., Akperov G. I., Alekseichik T. V. et al. *Soft Models of Management in Terms of Digital Transformation*. Rostov-on-Don : PEI HE SU (IMBL), 2019. 188 p.
3. Magerramov I. M. Decision Making Modeling in Fuzzy Criteria Conditions // *Intellectualnye resursy – regionalnomu razvitiyu*. 2021. Vol. 7, No. 1. P. 157–163. (In Russian).
4. Power BI Documentation. URL: <https://docs.microsoft.com/en-us/power-bi/> (accessed: 20.01.2022). (In Russian).
5. Google Data Studio. Connect and Visualize Your Data in Data Studio. URL: <https://developers.google.com/datastudio?hl=ru> (accessed: 20.11.2021).
6. Kramarov S. O., Khramov V. V., Belyaev A. V., Grebenyuk E. V. The Formalization of the Characteristics of Automatic Workplace (AWP) with the Use of Linguistic Variables (LV) // *IOP Conference Series: Materials Science and Engineering*, September 11–13, 2020. Rostov-on-Don, 2021. Vol. 1029. P. 012082. DOI 10.1088/1757-899X/1029/1/012082.
7. Chernyshev Yu. O., Khramov V. V. Osobennosti agregirovaniia kachestvennykh priznakov opornykh orientirov v sistemakh tekhnicheskogo zreniia // *Izvestiia TRTU*. 2001. No. 3 (21). P. 55. (In Russian).

8. Храмов В. В. Агрегирование информации как проблема личностной самоорганизации // Рос. психолог. журн. 2007. Т. 4, № 4. С. 9–21.
9. Крамаров С. О., Храмов В. В. Системно-инженерный подход к исследованиям сложных многомерных систем на основе мягких моделей // Интеллект. ресурсы – региональному развитию. 2018. Т. 4, № 1. С. 222–228.
10. Храмов В. В. Концепция обеспечения эффективности организационно-технических систем на основе бионико-интеллектуального подхода // Вестник Ростов. гос. ун-та путей сообщения. 2001. № 2. С. 138–141.
11. Khramov V. V. The Concept of Connectivity Model Measurements of Geographic and Information Space of the Region // Advances in Intelligent Systems and Computing. 2021. Vol. 1323 AISC. P. 553–558. DOI 10.1007/978-3-030-68004-6_72.
12. Голубенко Е. В., Гребенюк Е. В., Храмов В. В. Мягкие методы и модели эффективного управления образовательными системами // Актуальные проблемы и перспективы развития транспорта, промышленности и экономики России : сб. науч. тр. Ростов-на-Дону, 9–11 ноября 2020 г. Ростов-на-Дону : Ростов. гос. ун-т путей сообщения, 2020. С. 67–71.
8. Khramov V. V. Information Aggregating as a Problem of Personal Self-Organization // Russian Psychological Journal. 2007. Vol. 4, No. 4. P. 9–21. (In Russian).
9. Kramarov S. O., Khramov V. V. Sistemno-inzhenernyi podkhod k issledovaniyam slozhnykh mnogomernykh sistem na osnove miagkikh modelei // Intellectualnye resursy – regionalnomu razvitiu. 2018. Vol. 4, No. 1. P. 222–228. (In Russian).
10. Khramov V. V. Kontseptsiiia obespecheniia effektivnosti organizatsionno-tekhnicheskikh sistem na osnove bioniko-intellektualnogo podkhoda // Vestnik Rostov. gos. un-ta putei soobshcheniia. 2001. No. 2. P. 138–141. (In Russian).
11. Khramov V. V. The Concept of Connectivity Model Measurements of Geographic and Information Space of the Region // Advances in Intelligent Systems and Computing. 2021. Vol. 1323 AISC. P. 553–558. DOI 10.1007/978-3-030-68004-6_72.
12. Golubenko E. V., Grebenyuk E. V., Khramov V. V. Miagkie metody i modeli effektivnogo upravleniia obrazovatelnyimi sistemami // Aktualnye problemy i perspektivy razvitiia transporta, promyshlennosti i ekonomiki Rossii : Collection of scientific works. Rostov-on-Don, November 9–11, 2020. Rostov-on-Don : Rostov State Transport University, 2020. P. 67–71. (In Russian).

Информация об авторах

И. М. Магerrамов – аспирант.

Г. И. Акперов – директор по маркетингу и инновациям.

Information about the authors

I. M. Magerramov – Postgraduate.

G. I. Akperov – Director of Marketing and Innovations.