

ТЗ на парсинг сайта книжного магазина

Задача – спарсить информацию по книгам из книжного интернет-магазина и занести ее в таблицу ниже.

Формат таблицы



Книги.xlsx

Весь процесс

1. Берем все доступные книги из каталога.
2. Парсим информацию по книгам и заносим данные в таблицу.
3. Парсим информацию по остаткам в оффлайн магазинах.
4. Проверяем наличие книги на ВБ по ISBN
5. Проверяем наличие книги на ВБ по названию

1. Берем все доступные книги из каталога

Каталог открываем только по ссылкам ниже. Там в ссылках уже зашиты нужные категории и фильтры (года издания и статус наличия в оффлайн магазинах).

- <https://www.chitai-gorod.ru/catalog/books/hudozhestvennaya-literatura-110001?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/komiksy-110063?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/manga-110064?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/detskie-knigi-110090?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/knigi-dlya-podrostkov-110132?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/medicina-i-zdorove-110207?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>

[reWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023](https://www.chitai-gorod.ru/catalog/books/krasota-moda-ehetiket-110227?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023)

- <https://www.chitai-gorod.ru/catalog/books/krasota-moda-ehetiket-110227?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/kulinariya-110235?page=1&filters%5BonlyAvailableInCustomerCity%5D=1>
- <https://www.chitai-gorod.ru/catalog/books/dom-i-hobbi-110244?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/religiya-i-filosofiya-110307?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/psihologiya-110321?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/iskusstvo-110338?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/ehkonomika-menedzhment-marketing-110356?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>
- <https://www.chitai-gorod.ru/catalog/books/istoriya-obshchestvo-110388?page=1&filters%5BonlyAvailableInCustomerCity%5D=1&filters%5BliteratureWorkPublishingYearsMin%5D=2020&filters%5BliteratureWorkPublishingYearsMax%5D=2023>

На странице 16 книг, кол-во страниц зависит от ссылки. Парсим информацию по всем книгам на всех страницах.

2. Заходим во книги

Заходим в каждую книгу, например: <https://www.chitai-gorod.ru/product/instrukciya-k-zhizni-s-lyubovyu-ot-shkya-3005689?productShelf=&shelfIndex=0&productIndex=0>

Скриншот всей страницы ниже.

Там еще ниже продолжение)

себя новые интересные формулировки. Рекомендую.

Минусы

Минусов нет, только смех, юмор и веселье.

0

1

Показать больше отзывов

ОПИСАНИЕ И ХАРАКТЕРИСТИКИ

Российские комиксы

Читайте и помогайте!

Комиксы и манга

«Это самое тупое, что я читал». — Олег

«Очень скучно, еле до середины дошёл». — Виталий

«Книгу покупать не стала, слишком дорогая». — Светлана

«Почему так уродко нарисовано?» — Сергей

«Трогательная. Искромётная. Сумасшедшая. Эта книга переносит тебя в иной мир, из которого не хочется возвращаться. Уникальный юмор, художественное мастерство и поразительная эрудированность автора не смогут оставить равнодушным даже самого искушённого читателя. Это талант, граничащий с гениальностью». — Ева Морозова

ID товара

3005689

Издательство

ЭКСМО

Год издания

2023

ISBN

978-5-04-188079-8

Количество страниц

104

Размер

22,7×17,5×1,3

Тип обложки

Твёрдый переплёт

Тираж

8000

Вес, г

410

Возрастные ограничения

16+

1199 Р

994 Р

+ до 149 бонусов

Купить

В наличии

8 магазинов сети, бесплатно завтра

Адреса магазинов

8 пункты выдачи, 140 Р завтра

Пункты выдачи

Доставка курьером, 225 Р завтра

Варианты доставки

Экспресс-доставка

Цена товара — 1199 Р

За 3 часа по Москве, 800 Р

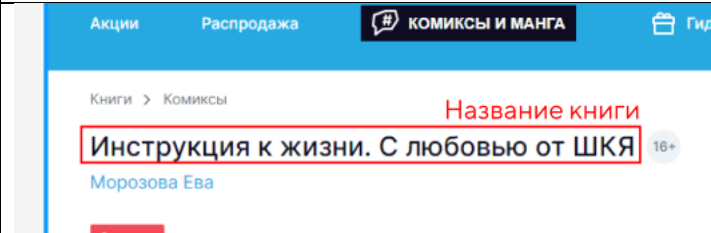
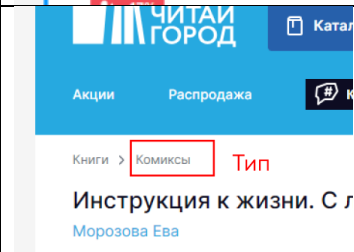
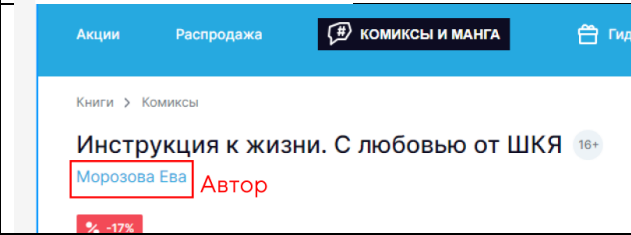
Сегодня

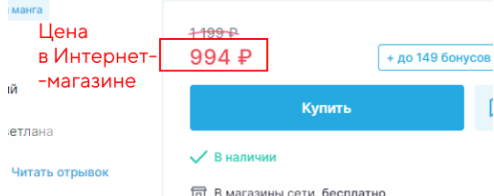
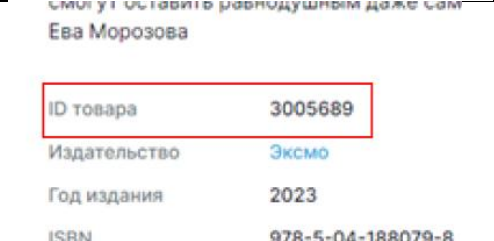
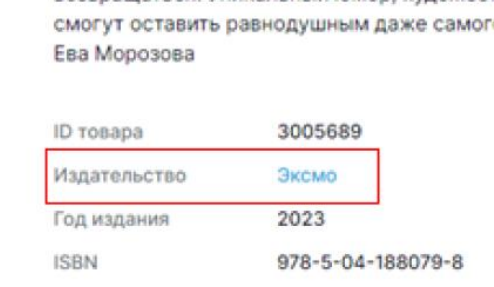
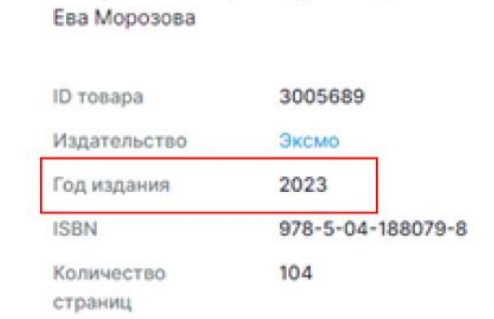
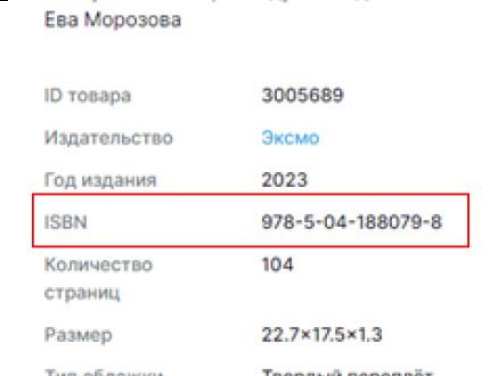
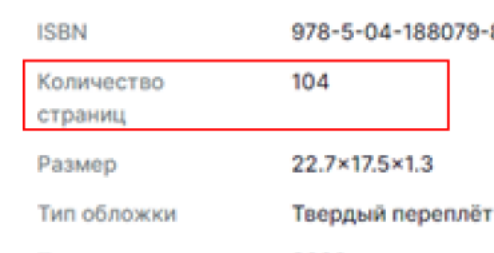
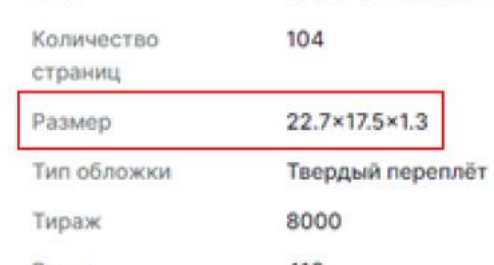
Сформировать экспресс-доставку

В наличии в 76 магазинах сети

Там еще ниже есть рекомендации и подвал, но нам это уже не нужно.

Данные складываем в таблицу. Ниже указал, откуда брать информацию для каждого поля таблицы.

Название поля	Откуда берем	Пример заполнения
Название книги		Инструкция к жизни. С любовью от ШКЯ
Тип		Комиксы
Автор		Морозова Ева

Цена в интернет- магазине		994
ID товара		3005689
Издательст во		Эксмо
Год издания		2023
ISBN		978-5-04-188079-8
Кол-во страниц		104
Размер		22.7x17.5x1.3

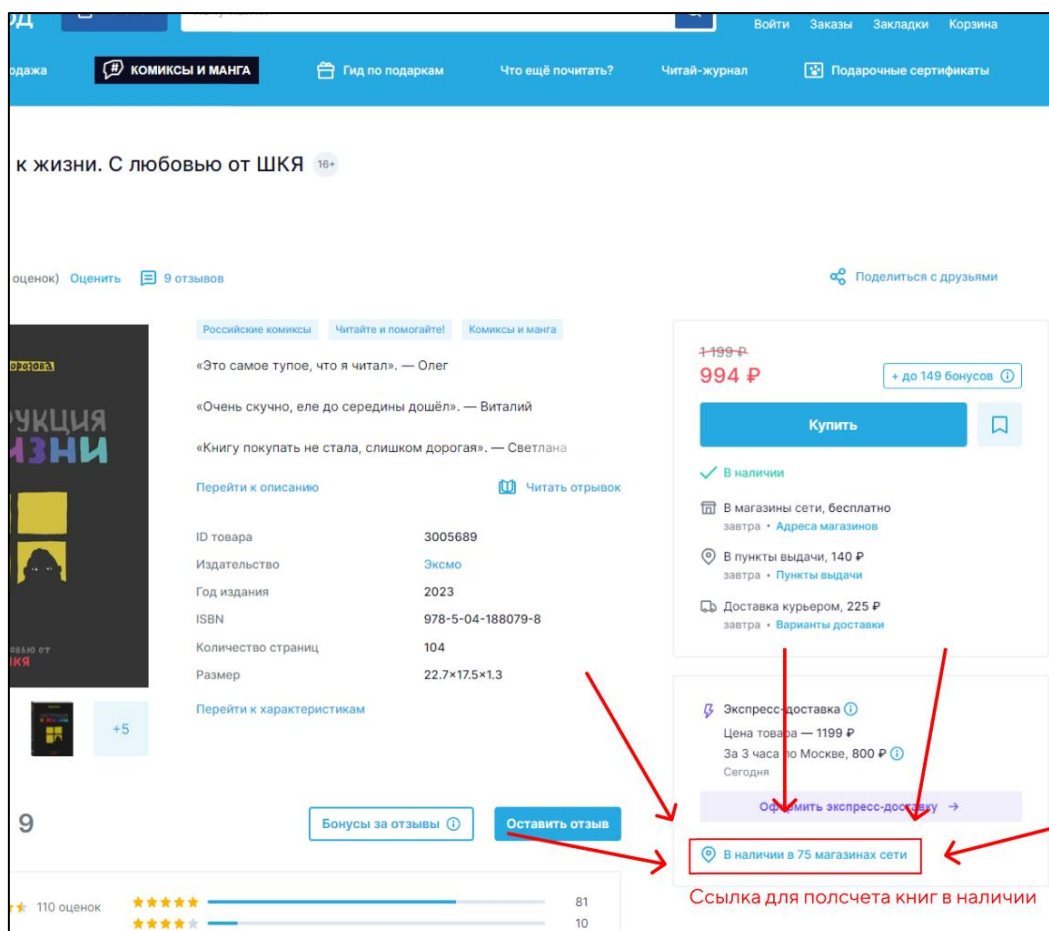
Тип обложки	<div> <div>страниц</div> <div>Размер 22.7×17.5×1.3</div> <div>Тип обложки Твердый переплёт</div> <div>Тираж 8000</div> <div>Вес, г 410</div> <div>Размер 22.7×17.5×1.3</div> <div>Тип обложки Твердый переплёт</div> <div>Тираж 8000</div> <div>Вес, г 410</div> </div>	Твердый переплёт
Тираж	<div> <div>Тип обложки Твердый переплёт</div> <div>Тираж 8000</div> <div>Вес, г 410</div> </div>	8000
Вес, г	<div> <div>Тираж 8000</div> <div>Вес, г 410</div> <div>Возрастные 16+</div> </div>	410
Возрастное ограничение	<div> <div>Тираж 8000</div> <div>Вес, г 410</div> <div>Возрастные ограничения 16+</div> </div>	16+
Средний бал	<div>  <h3>ОТЗЫВЫ 9</h3> <p>Средний бал</p> <p>4.3 ★★★★★ 110 оценок ★★★★★</p> </div>	4.3
Количество оценок, шт	<div> <h3>ОТЗЫВЫ 9</h3> <p>Кол-во оценок</p> <p>4.3 ★★★★★ 110 оценок ★★★★★</p> </div>	110
Статус наличия в Интернет магазине	<div> <p>994 ₽ + до 149 бонусов</p> <p>Купить</p> <p>✓ В наличии</p> <p>В магазины сети, бесплатно завтра • Адреса магазинов</p> <p>В пункты выдачи, 140 ₽</p> <p>Статус наличия в Интернет-магазине</p> </div>	В наличии

<p>Наличие в магазинах Москвы</p>	<div><div>купить</div><div><div><div>✓ В наличии</div><div><div>В магазины сети, бесплатно</div><div>завтра • Адреса магазинов</div></div><div><div>В пункты выдачи, 140 ₽</div><div>завтра • Пункты выдачи</div></div><div><div>Доставка курьером, 225 ₽</div><div>завтра • Варианты доставки</div></div></div></div><div><div>Экспресс-доставка</div><div>Цена товара — 1199 ₽</div><div>За 3 часа по Москве, 800 ₽</div><div>Сегодня</div><div>Оформить экспресс-доставку →</div><div><div>В наличии в 75 магазинах сети</div></div></div></div>	<p>75</p>
<p>Описание</p>	<div><div>ОПИСАНИЕ И ХАРАКТЕРИСТИКИ</div><div><div>Российские комиксы</div><div>Читайте и помогайте!</div><div>Комиксы и манга</div><div>Описание</div></div><div><div>«Это самое тупое, что я читал». — Олег</div><div>«Очень скучно, еле до середины дошёл». — Виталий</div><div>«Книгу покупать не стала, слишком дорогая». — Светлана</div><div>«Почему так уродско нарисовано?» — Сергей</div><div>«Трогательная. Искромётная. Сумасшедшая. Эта книга переносит тебя в иной мир, из которого не хочется возвращаться. Уникальный юмор, художественное мастерство и поразительная эрудированность автора не смогут оставить равнодушным даже самого искушённого читателя. Это талант, граничащий с гениальностью». — Ева Морозова</div></div><div><div>ID товара</div><div>3005689</div><div>Издательство</div><div>ЭКСМО</div><div>Год издания</div><div>2022</div></div></div>	<div><div>«Это самое тупое, что я читал». — Олег</div><div>«Очень скучно, еле до середины дошёл». — Виталий</div><div>«Книгу покупать не стала, слишком дорогая». — Светлана</div><div>«Почему так уродско нарисовано?» — Сергей</div><div>«Трогательная. Искромётная. Сумасшедшая. Эта книга переносит тебя в иной мир, из которого не хочется возвращаться. Уникальный юмор, художественное мастерство и поразительная эрудированность автора не смогут</div></div>

		оставить равнодушным даже самого искущённого читателя. Это талант, граничащий с гениальностью». — Ева Морозова
--	--	--

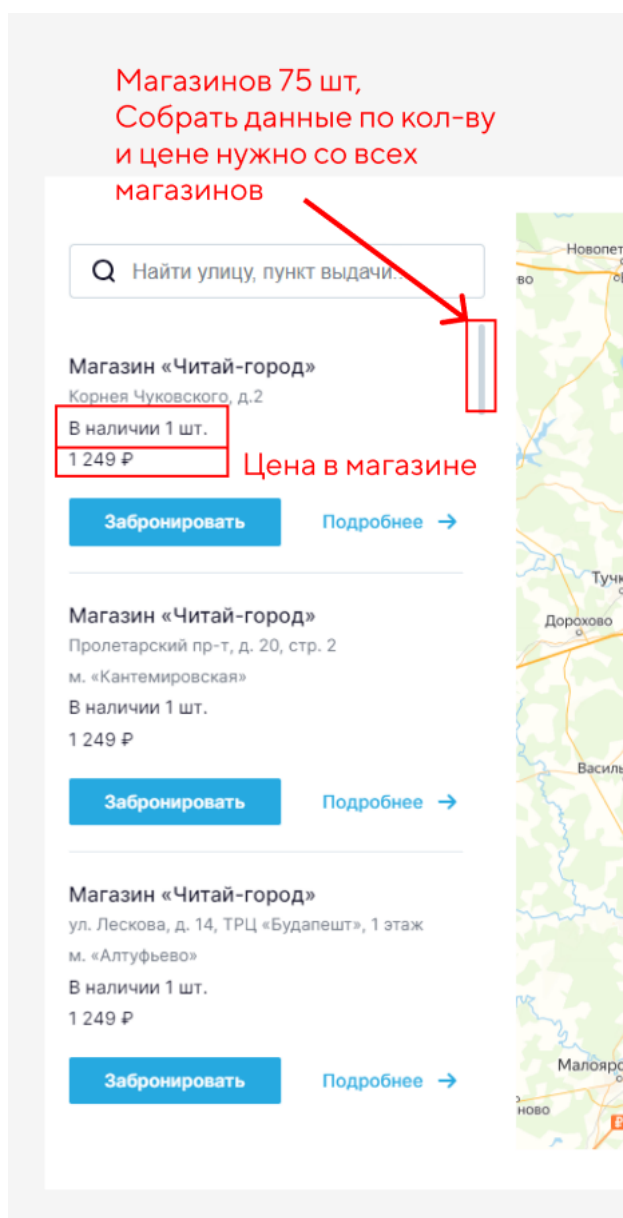
3. Парсим информацию по остаткам и ценам в оффлайн магазинах

Чтобы посчитать, сколько штук доступно в магазинах Москвы, необходимо перейти по ссылке, указанной на картинке ниже



Вот пример ссылки: <https://www.chitai-gorod.ru/availability/product/instrukciya-k-zhizni-s-lyubovyu-ot-shkya-3005689>

Далее необходимо суммировать количество книг доступные в каждом магазине и рассчитать среднюю, максимальную и минимальную цену, по которой книга доступна в магазине.



Данные заносим в таблицу:

Название поля	Пример
Средняя цена	1249
Минимальная цена	1249
Максимальная цена	1249
Кол-во штук по всем магазинам	148

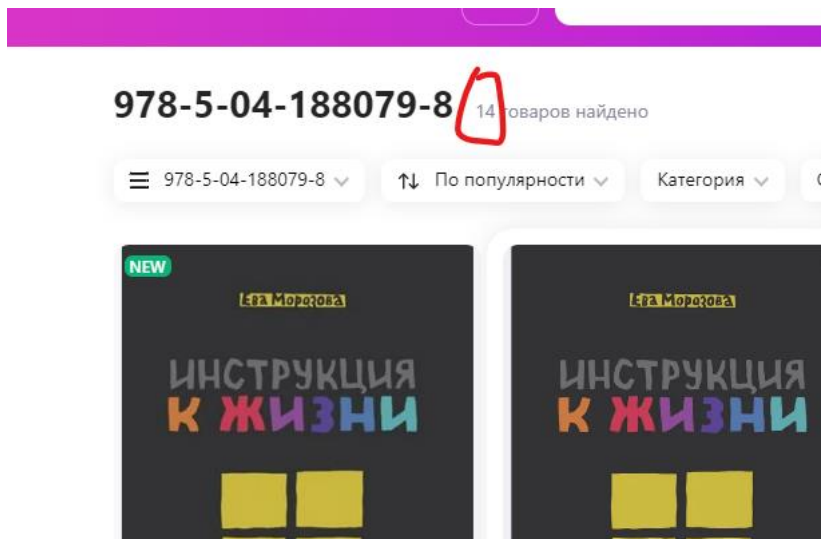
4. Проверяем наличие книги на ВБ по ISBN

Для этого открываем поисковый запрос ВБ и подставляем в него ISBN

Например, для нашей книги:

<https://www.wildberries.ru/catalog/0/search.aspx?search=978-5-04-188079-8>

Берем и заносим в таблицу кол-во найденного товара



Поля в таблице:

Название поля	Пример
Наличие книги на ВБ по ISBN	14

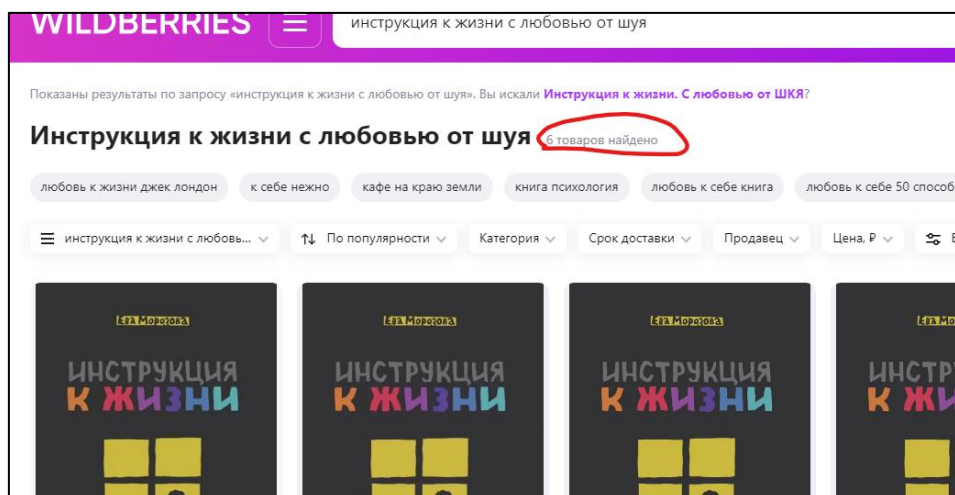
5. Проверяем наличие книги на ВБ по названию

Для этого открываем поисковый запрос ВБ и подставляем в него ISBN

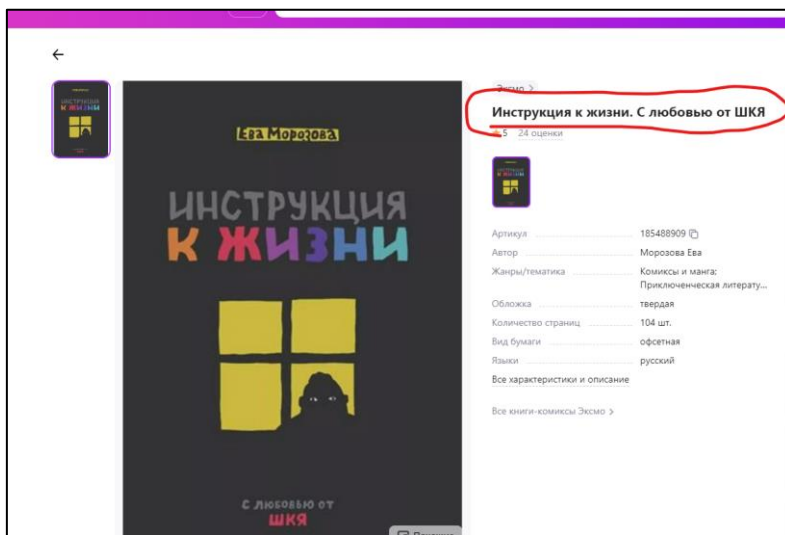
Например, для нашей книги: [https://www.wildberries.ru/catalog/0/search.aspx?search=Инструкция к жизни. С любовью от ШКЯ](https://www.wildberries.ru/catalog/0/search.aspx?search=Инструкция+к+жизни.+С+любовью+от+ШКЯ)

Далее действуем так:

- Также, как и по ISBN фиксируем в таблице кол-во найденных товаров



- Если кол-во найденных товаров меньше 40, то заходим во все найденные карточки. Если больше 40, то заходим в первые 40.
- Сверяем первые 62 символа (если меньше 62, то берем сколько есть) названия на ВБ с первыми 62 символами названия, по которому делали поиск по ВБ



- Если хотя бы у одной карточки из 40 название совпадает с нашим названием, то
 - В таблице проставляем статус «Есть».
 - Проставляем позицию первой совпавшей карточки (например, у первых 10 карточек название отличалось, а у 11 совпало, значит в таблицу проставляем 11)
 -
- Если ни одного совпадения, то проставляем в таблице статус «Нет»

Поля в таблице:

Название поля	Пример
Результат поиска на ВБ по названию	6
Наличие книги на ВБ по названию	Есть
Позиция первой карточки ВБ, где совпали названия	1
Количество карточек, где было совпадение по названию	5