**Создание робота по считыванию цен с сайтов товаров**

**1. Базовые понятия**

В данном разделе определяется ряд базовых понятий, которые будут использоваться в дальнейшем (рисунок 1.1):

* *Рабочая область* – центральное окно в программе PIX Studio после создания проекта
* *Окно* *активностей* – левое окно в программе PIX Studio после создания проекта
* *Активности* – базовые функции, из которых конструируется робот
* *Группа активностей* – активности, объединённые единой идей своего функционала. В дальнейшем для ясности активности будут именоваться следующим образом «Группа активностей/Название активности»
* *Использовать/применить активность* – перетащить активность из окна активностей в рабочую область. Для выполнения данного действия требуется нажать ЛКМ на активность и, зажав ЛКМ, перенести курсор в рабочую область, а затем отпустить ЛКМ

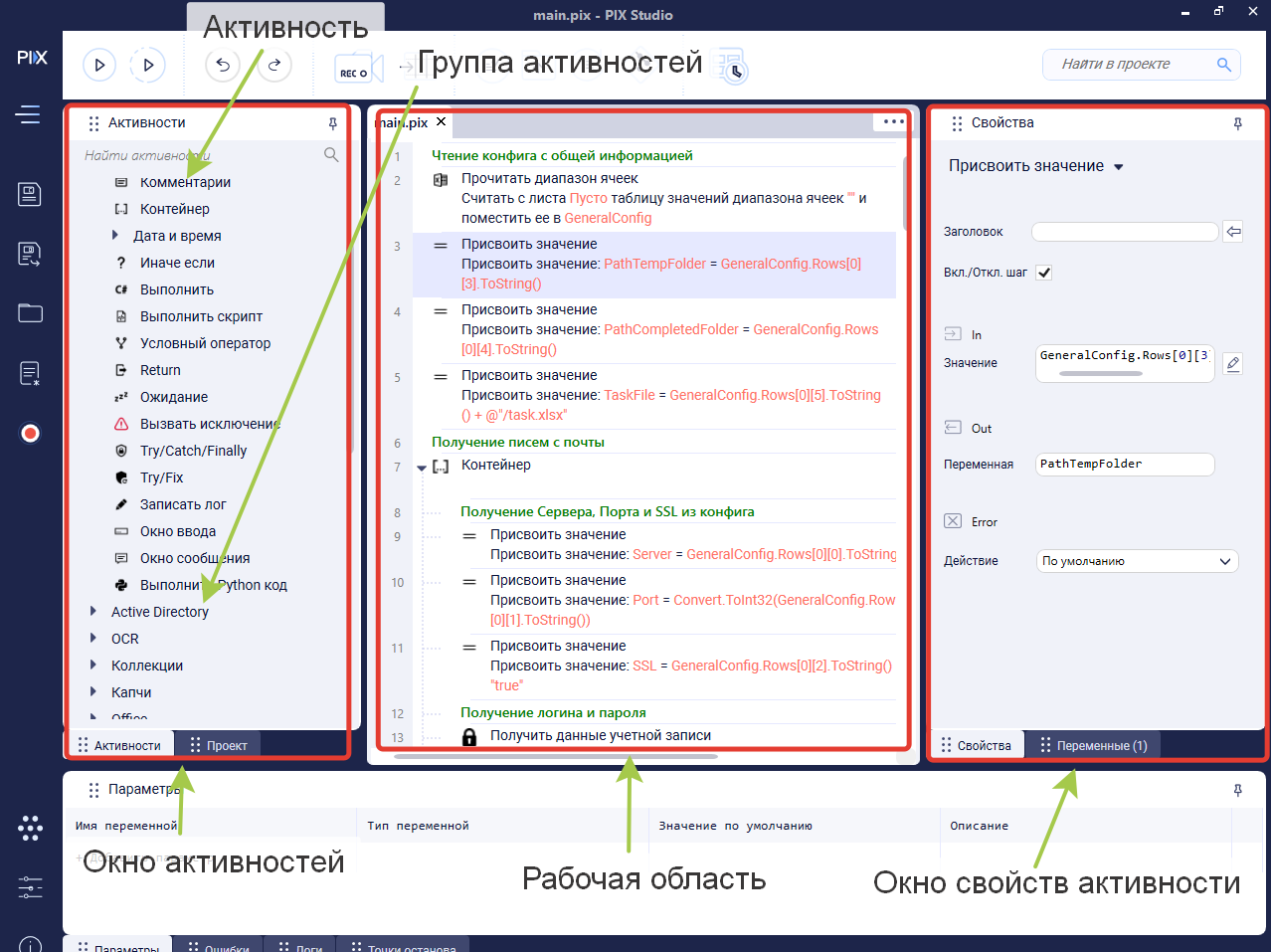


Рисунок 1.1 Окно редактора PIX Studio

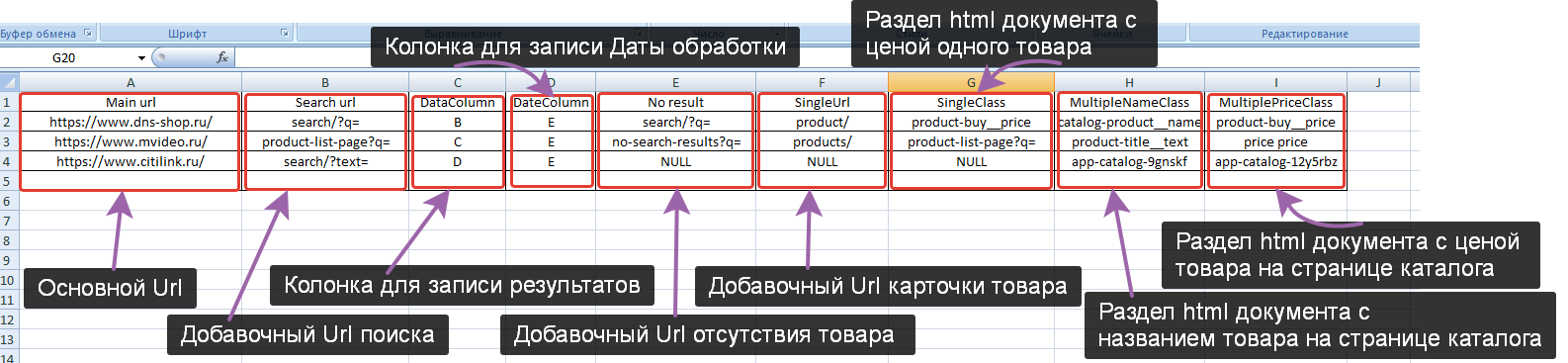
**2. Постановка задачи**

Робот должен прочитать содержимое шаблона excel документа и по всем полученным названиям товаров, проверить их наличие на различных сайтах и в случае его наличия взять цену товара и записать в excel документ.

**3. Предварительные действия**

Прежде чем приступить к созданию робота, желательно выполнить ряд подготовительных действий. В первую очередь, создаётся дополнительный excel документ, который в дальнейшем будет именоваться, как «*конфиг*». В данные документ заносится информация, которая необходима для работы робота, но может быть изменена сторонними действиями. Для данной задачи конфиг может содержать следующую информацию (Рисунок 3.1):

* URL сайтов, и дополнительные URL для определения наличия товара и поиска товара через URL.
* Наименование колонок, куда записывать считанные цены.



3.1. Пример конфига для робота.

Помимо конфига необходимо убедится в наличие шаблона excel таблицы, куда будут занесены считанные цены. Пример такой таблицы можно увидеть на рисунке 3.2.

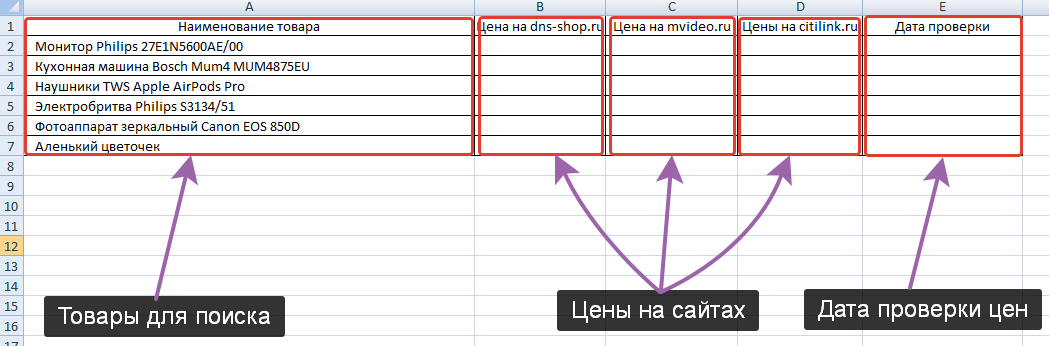


Рисунок 3.2. Пример шаблона excel таблицы.

**4. Работа с PIX Studio**

Теперь необходимо запустить PIX Studio и создать новый проект. Для этого сначала необходимо нажать на панели слева на иконку блокнота со звёздочкой в левом нижнем углу. В появившемся окне выбирается пункт проект (Рисунок 4.1). PIX Studio попросит ввести название проекта и его расположение.

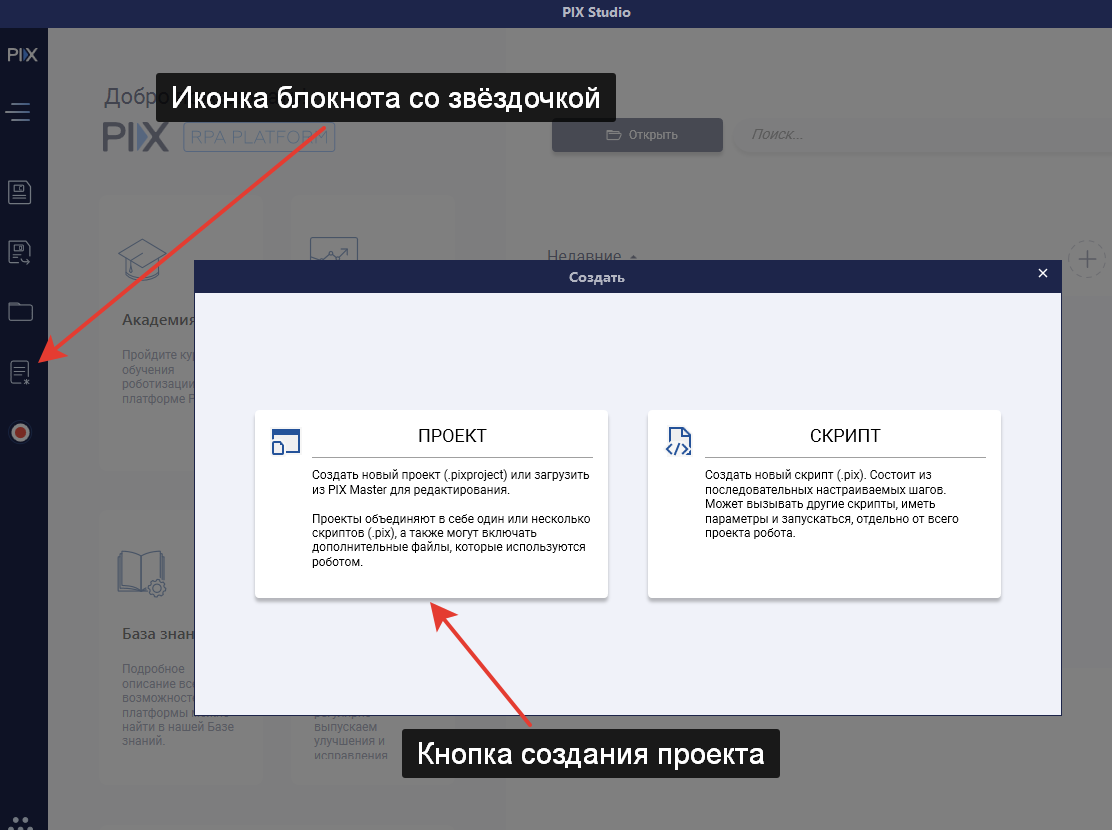


Рисунок 4.1. Создание проекта

**4.1. Чтение входных данных и создание глобальных переменных**

В первую очередь с помощью активности *«Базовые/Присвоить значение»* создаются переменный с путями к необходимым файлам, конфигу и т.д.

Затем с помощью активности *«Office/Excel/Прочитать диапазон ячеек»* читается конфиг и excel шаблон файла с товарами. В свойствах активности заполняются: *Лист* – лист, с которого необходимо прочитать данные, *Диапазон* – диапазон, который нужно прочитать, в случае если нужны все непустые ячейки указывает «””», *С заголовками* – ставится галочка, если в файле есть заголовки, *Путь к фалу* – путь нужному файлу, *Таблица* – переменная, в которую будет записана таблица с результатом.

Также необходимо копировать файл шаблона в папку с результатом активностью *«Файлы/Копировать файл/папку»*. В свойствах указывается: *Путь откуда* – путь к шаблону, *Путь куда* – путь куда должен быть сохранён результирующий файл с названием.

А также активностью *«Базовые/Дата и время/Получить дату и время»* получается текущая дата и время. А затем с помощью активности *«Базовые/Дата и время/Дата и время в виде строки»* дата преобразуется в строку нужного формата. В свойствах указывается: *Дата и время* – переменная, в которую была сохранена текущая дата, *Формат* – формат, в котором нужно записать дату, *Результат* – переменная, в которую сохраняется строка с результатом. (Рисунок 4.1.1)

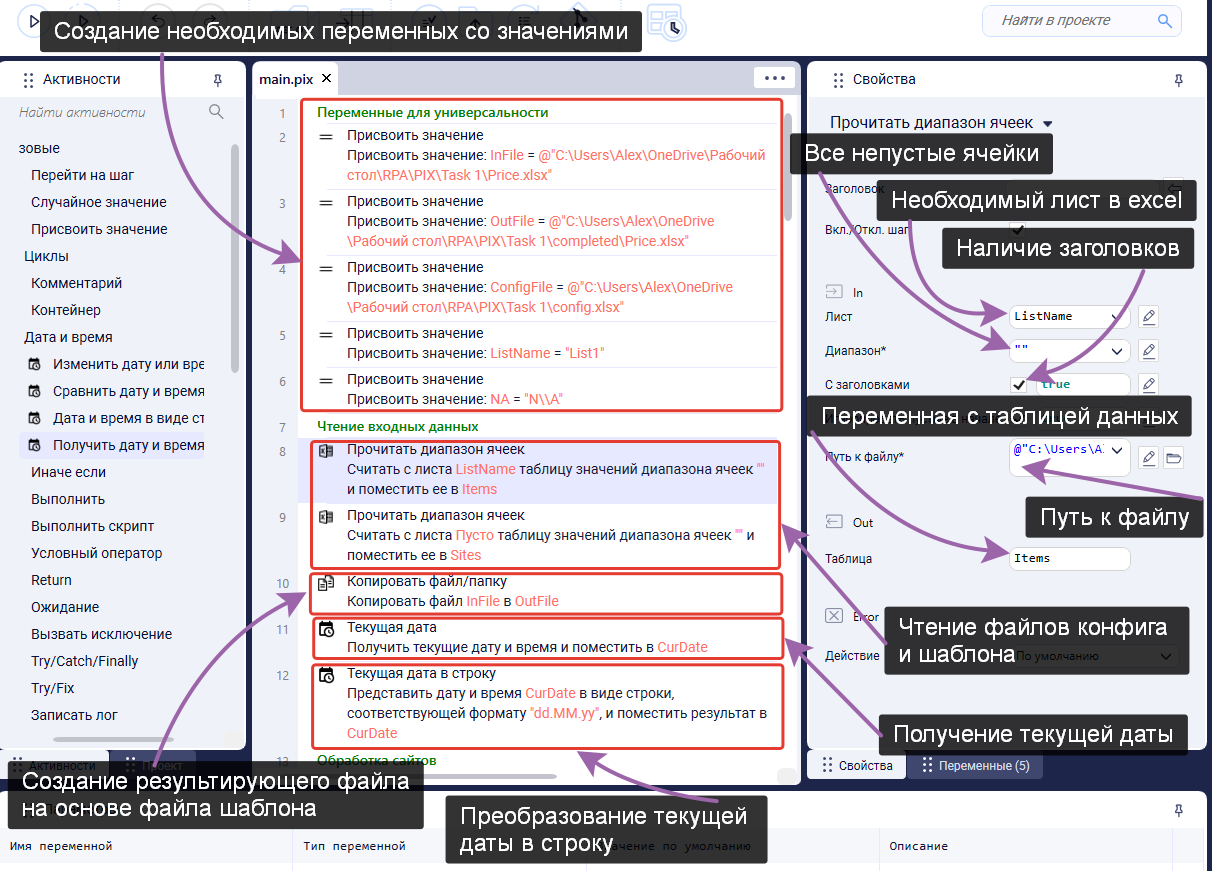


Рисунок 4.1.1. Чтение входных данных

**4.2. Обработка сайтов**

С помощью активности *«Базовые/Циклы/Цикл для каждого»* запускается цикл по строкам таблицы файла конфига, что то же самое, что и цикл по сайтам.

В начале цикла все значения ячеек строки таблицы записываются в отдельные переменные.

А также с помощью активности *«Web/Запусть Google Chrome»* запускается браузер. В свойствах активности указывает: *URL* – переменная со значением основного URL для сайта, *Браузер* – переменная, в которую будет сохранено соединение данного браузера. (Рисунок 4.2.1)

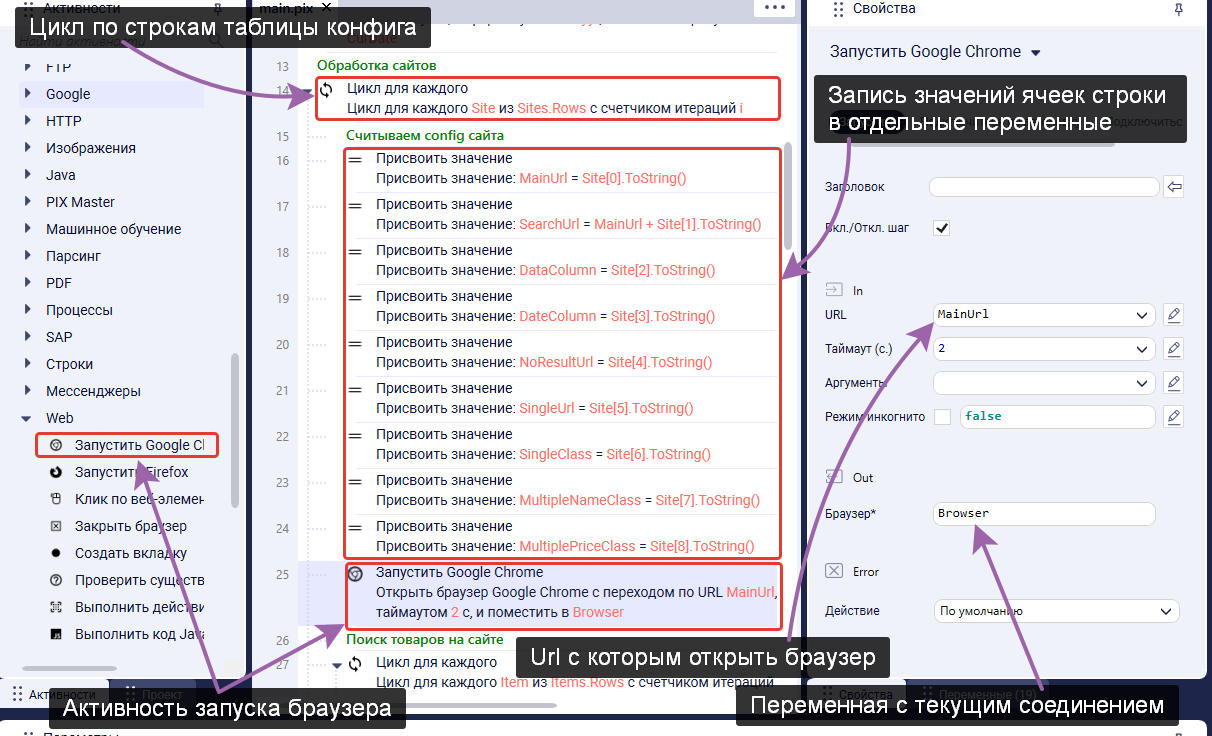


Рисунок 4.2.1. Подготовительная работа с сайтом

Затем активностью *«Базовые/Циклы/Цикл для каждого»* запускается цикл по необходимым товарам.

Первым делом, активностью *«Базовые/Присвоить значение»* сохраняется название товара в отдельную переменную.

После с помощью активности «*Web/Перейти по URL»* осуществляется переход на страницу поиска конкретного товара. В свойствах активности заполняются: *URL* – URL поиска конкретного товара, *Браузер* – переменная с текущим соединением с браузером.

И с помощью активности *«Базовые/Присвоить значение»* и C# функции *Соединение.URL* текущий URL сохраняется в отдельную переменную.

Следующий шагом активностью *«Базовые/Условный оператор»* и C# функцией *Строка.Contains(Подстрока)* происходит сверка текущего URL с URL'ом отсутствия товара. Если проверка успешная, то в соответствующую ячейку результирующего файла записывается информация об отсутствие товара активностью *«Office/Excel/Записать в ячейку»*. В свойствах активности указывается: *Лист* – лист excel файла, на котором нужно записать информацию, *Ячейка* – ячейка, в которую нужно записать данные, *Значение* – данные, которые надо записать, *Путь к файлу* – путь к результирующему файлу. (Рисунок 4.2.2)

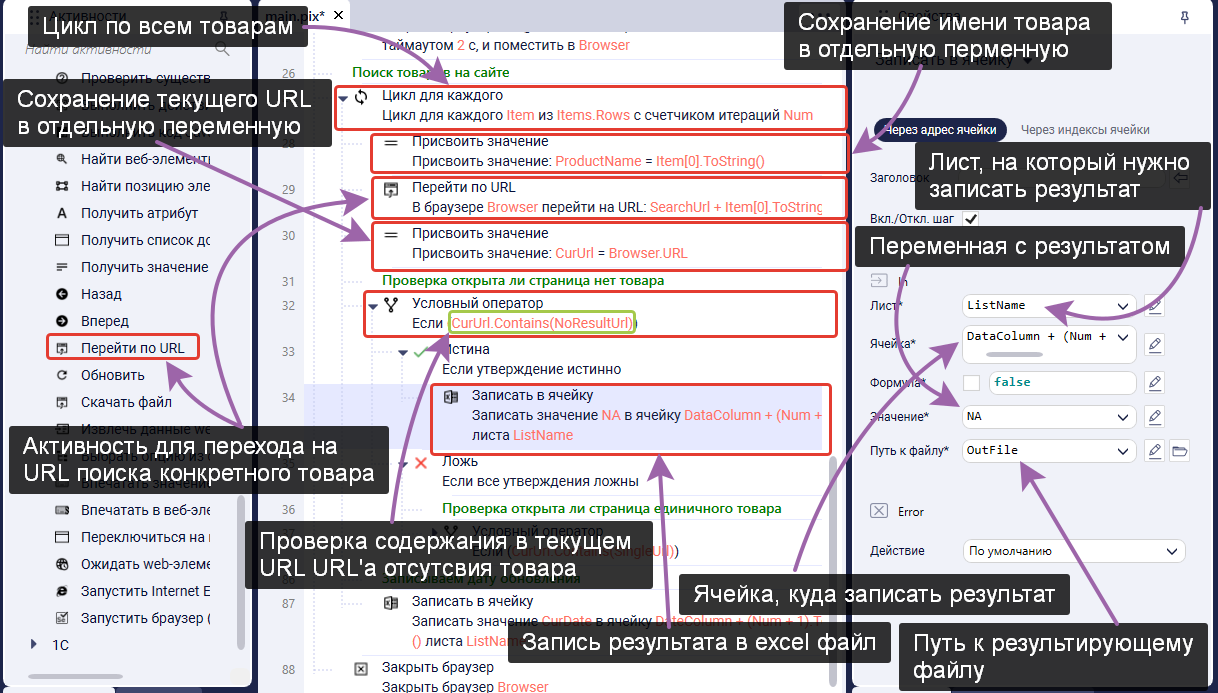


Рисунок 4.2.2. Проверка на отсутствие товара

Если результат проверки отрицательный, то активностью *«Базовые/Условный оператор»* и C# функцией *Строка.Contains(Подстрока)* осуществляется проверка URL на соответствие странице конкретного товара.

Если проверка прошла успешно, то активностью *«Web/Получить значение веб-элемента»* получается цена товара. В свойствах указывается: *Искать по* – указывается Class, так как поиск по классу на html-странице, *Параметр поиска* – элемент, который ищется, *Браузер* – текущее соединение с браузером, *Результат* – переменная, куда будет сохранён результа, *Таймаут* – время поиск веб-элемента.

После получение цены, она записывается в excel файл активностью *«Office/Excel/Записать в ячейку».* (Рисунок 4.2.3)

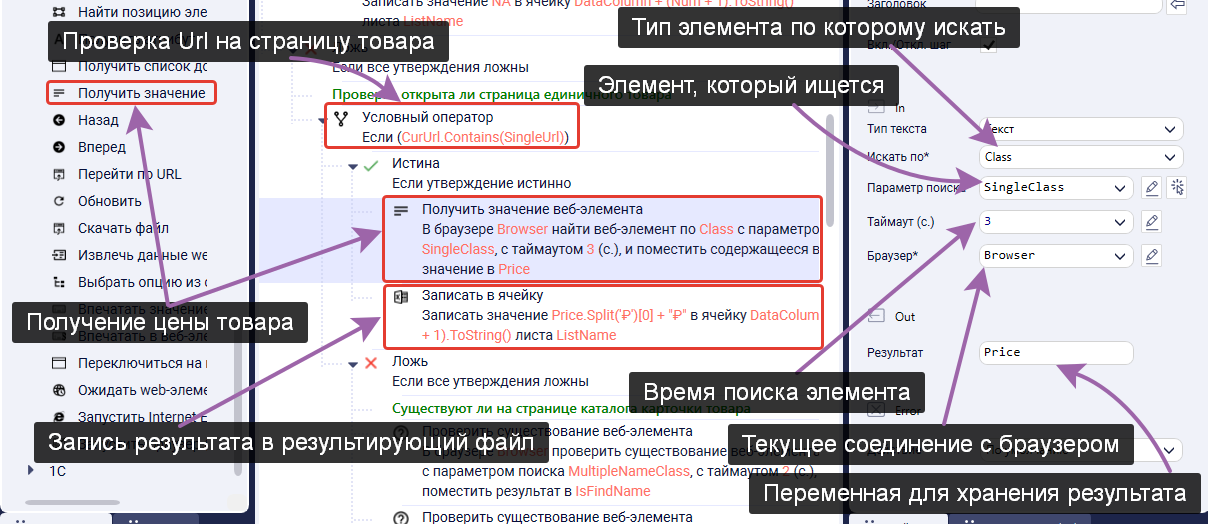


Рисунок 4.2.3. Поиск цены на странице товара.

Если же и эта проверка не прошла, то активностью *«Web/Проверить существование веб-элемента»* проверяется наличие хоть одного товара в каталоге с ценой и названием. В свойствах активности указывается: *Искать по* – ставится Class, *Параметр поиска* – элемент, который ищется, *Таймаут* – время поиска, *Браузер* –текущее соединение с браузером, *Результат* – булева переменная результата.

Если найдена и цена и название кого-то товара на странице каталога, то активностью *«Web/Найти веб-элементы»* забираются все ценники и названия товаров со страницы. Также активностью *«Базовые/Присвоить значение»* создаются переменные для поиска наилучшей цены. (Рисунок 4.2.4)

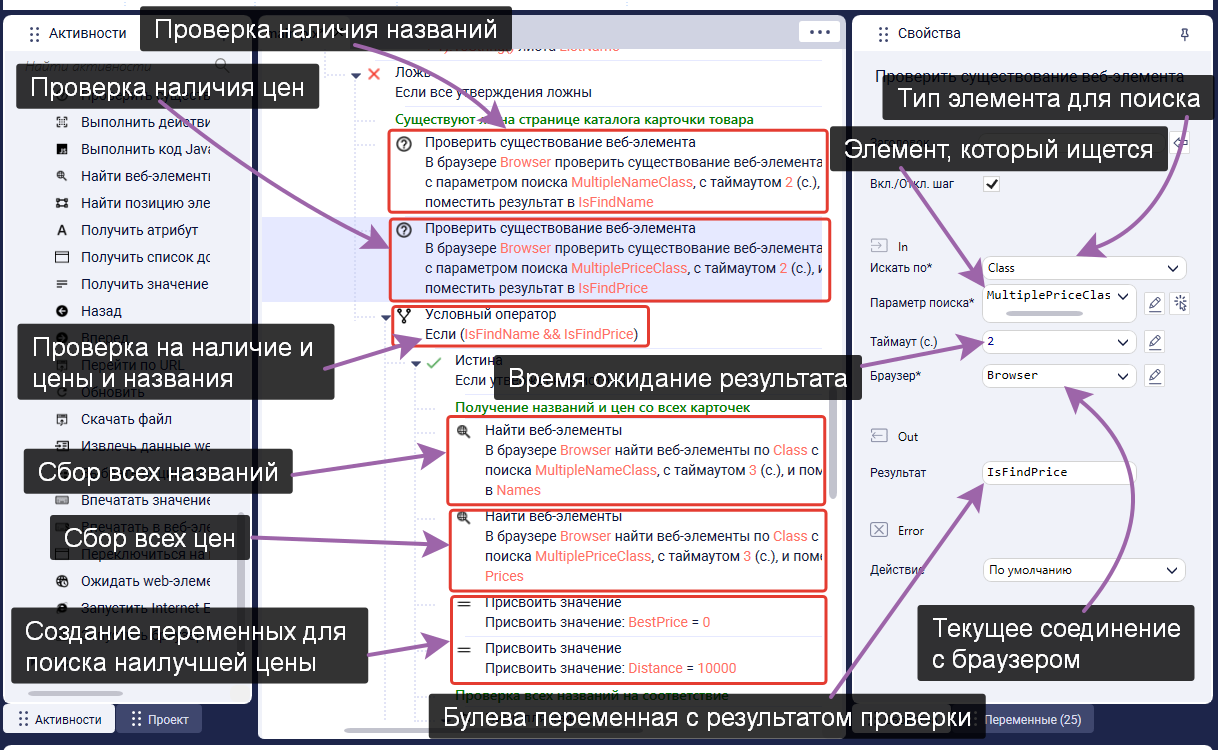


Рисунок 4.2.4. Поиск всех цен и названий

Затем запускается цикл по всем полученным названиям активностью *«Базовые/Циклы/Цикл для каждого»*. Активностями *«Базовые/Циклы/Цикл для каждого»* и *«Базовые/Условный оператор»* и C# функцией *Строка.Contains(Подстрока)* проверяется наличие всех слов из интересующего товара в считанном товаре.

Если все слова присутствуют, то активностью *«Строки/Расстояние Левенштейна»* считается расстояние Левенштейна и если оно меньше сохранённого значения, то сохранённое значение перезаписывается. (Рисунок 4.2.5)

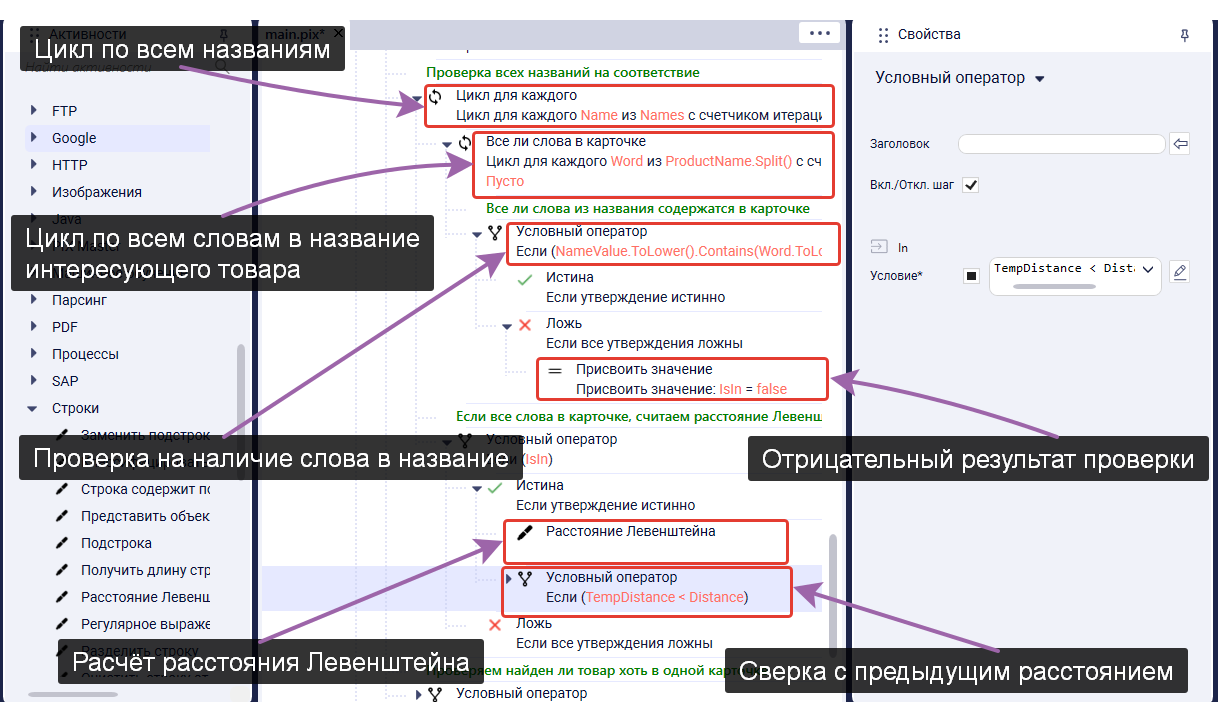


Рисунок 4.2.5. Проверка названия на соответствие

Затем с помощью активности *«Базовые/Условный оператор»* проверяется был ли найден хоть один товар. Если ответ положительный, то цена записывается в соответствующую ячейку excel документа с помощью активности *«Office/Excel/Записать в ячейку»*.

В случаях, если какая-то из раннее перечисленных проверок не прошла, то активностью *«Office/Excel/Записать в ячейку»* в соотвествующую ячейку записывается информация об отсутствие товара.

В конце цикла по всем товарам активностью *«Web/Закрыть браузер»* закрывается текущий браузер.

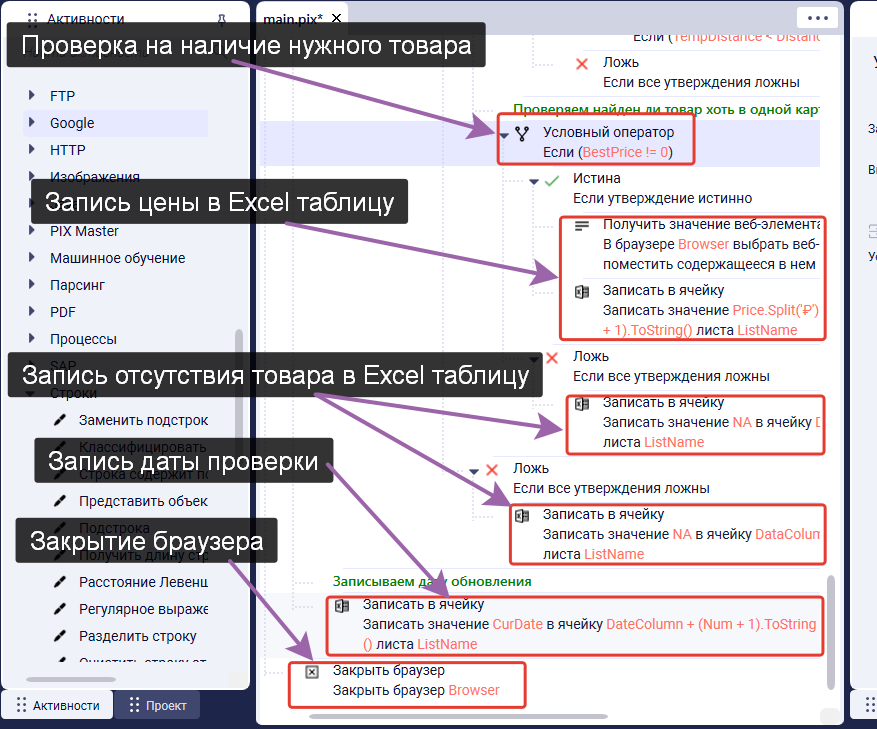


Рисунок 4.2.6. Запись данных в excel документ

**Литература**

1. <https://knowledgebase.pixrpa.ru/actions/Base/DateTimeToString>
2. <https://knowledgebase.pixrpa.ru/actions/Excel/ReadRange>
3. <https://knowledgebase.pixrpa.ru/actions/Files/CopyFileCatalog>
4. <https://knowledgebase.pixrpa.ru/actions/Base/GetDateTime>
5. <https://knowledgebase.pixrpa.ru/actions/Web/LaunchChrome>
6. <https://knowledgebase.pixrpa.ru/actions/Web/NavigateURL>
7. <https://knowledgebase.pixrpa.ru/actions/Web/ElementExist>
8. <https://knowledgebase.pixrpa.ru/actions/Web/FindElement>
9. <https://knowledgebase.pixrpa.ru/actions/Web/GetText>
10. <https://knowledgebase.pixrpa.ru/actions/Base/LoopForEach>
11. <https://knowledgebase.pixrpa.ru/actions/Base/Assign>
12. <https://knowledgebase.pixrpa.ru/actions/Strings/Levenshtein>
13. <https://knowledgebase.pixrpa.ru/actions/Base/If>
14. <https://knowledgebase.pixrpa.ru/actions/Excel/WriteCell>
15. https://knowledgebase.pixrpa.ru/actions/Web/CloseBrowser