

# Парсинг товаров с сайта средствами 1С 8

Взято тут: <http://oparin.info/2011/08/02/парсинг-товаров-с-сайта-средствами-1с-8>

В этой статье я постараюсь описать процесс парсинга сайтов средствами 1С с примером. Это статья не является инструкцией к применению, а лишь демонстрирует возможности 1С.

Что мы имеем?

- Сайт в интернете, на котором располагается список товаров. В моем случае — это интернет магазин салона «Связной»
- Понимание основ сайтостроения... хотя бы знание HTML тегов
- Умение кодить в 1С 8

Все вышеперечисленные пункты в арсенале? Тогда читаем далее

Для начала парсинга стоит определиться с тем что мы хотим спарсить и какая у нас будет иерархия. В моем случае это категория сотовых телефонов. Верхний уровень иерархии будет производители. Почему именно так? Потому что я так захотел. Вы же вправе использовать любую иерархию.

Далее нам будут интересны такие поля как: Наименование, Цена, Картинка и Описание... ну и пожалуй захватим операционную систему, чтобы пример получился более наглядным.

## Создаем внешнюю обработку

Те, кто не знают как это сделать — дальше могут не читать.

- Создаем форму обработки с командной панелью снизу и сверху (они могут быть полезными)
- Размещаем на ней Панель и обзываем первую страницу «СамСайт»
- Кладем на страницу «СамСайт» ПолеHTMLДокумента и обзываем его к примеру «Сайт»
- Переименовываем кнопку «Выполнить», которая находится на нижней панели в «Загрузить сайт»
- Описываем процедуру нажатия на эту кнопку так:

```
ЭлементыФормы.Сайт.Перейти("http://www.svyaznoy.ru/catalog/phone/224");  
/Категория с мобильными телефонами
```

Проверяем работу нашей обработки. У меня появился сайт связного. А у Вас?

Дальше сложнее. Все еще хочешь парсить сайты? Тогда читай

Сам парсинг сайта заключается в обходе всех элементов загруженной страницы, выдергивания необходимой информации и запихивания их в табличную часть. Для это:

- Создадим табличную часть «Производители» с реквизитами «Отметка» (Булево), «Наименование» (Строка 100) и «Ссылка» (Строка 300).

- Добавляем еще одну страницу на панели и обзываем ее «Производители»
- Размещаем на этой странице одноименную табличную часть
- Добавляем на нижней панели кнопку «Заполнить производителей» с кодом:

```
Для Каждого Стр из ЭлементыФормы.Сайт.Документ.body.all Цикл
  Если Стр.tagName = "H1" и Стр.innerHTML = "Производители" Тогда
    Для Каждого опСтр из Стр.nextSibling.children Цикл
      новСтр = Производители.Добавить();
      новСтр.Наименование = опСтр.innerHTML;
      новСтр.Ссылка = опСтр.firstChild.href;
    КонецЦикла;
  Возврат;
КонецЕсли;
КонецЦикла;
```

Здесь напрашиваются небольшие пояснения:

- tagName — имя HTML тега в HTML документе
- nextSibling — следующий элемент HTML документа от текущего
- children — список дочерних элементов
- firstChild — первый дочерний элемент от текущего

Проверяю. При проверке важно, находиться на странице «СамСайт», чтобы заполнять производителей

Производители заполнены. Теперь к самим телефонам

- Создаем табличную часть «Товары» с реквизитами «Производитель» (Строка 100), «Наименование» (Строка 100), «Цена» (Число 10,2), «Картинка» (Строка 300), «Описание» (Строка Неограниченная), «ОС» (строка 100), «Ссылка» (Строка,300)
- Добавляем еще одну страницу на панели и обзываем ее «Товары»
- Размещаем на этой странице одноименную табличную часть
- Добавляем на нижней панели кнопку «Заполнить товары» с кодом:

```
Для Каждого Стр из Производители Цикл
  //Если отметку сняли - то не трогаем этого производителя
  Если Не Стр.Отметка Тогда
    Продолжить;
  КонецЕсли;
  Форма = ПолучитьФорму("ФормаТоваров");
  Форма.ЭлементыФормы.Сайт.Перейти(Стр.Ссылка);
  Форма.ТекущийПроизводитель = Стр;
  Форма.ОткрытьМодально();
КонецЦикла;
```

- Создаем форму обработки «ФормаТоваров»
- Кладем на «ФормаТоваров» ПолеHTMLДокумента и называем его «Сайт»
- На событие ДокументСформирован у ПоляHTMLДокумента пишем код:

```
Если ЭлементыФормы.Сайт.Документ.body.all.length>1 Тогда
  ГрузимТовары();
```

КонецЕсли;

- Создаем переменную в модуле формы

```
перем ТекущийПроизводитель Экспорт;
```

- Создаем процедуру ГрузимТовары():

```
Процедура ГрузимТовары()
    Для Каждого Стр из ЭлементыФормы.Сайт.Документ.body.all Цикл
        Если Стр.className = "ct_desc cleared" Тогда
            новСтр = Товары.Добавить();
            Для Каждого опСтр из Стр.children Цикл
                Если опСтр.className = "pic_and_comp" Тогда
                    новСтр.Картинка =
СтрЗаменить(Сред(опСтр.firstChild.style.backgroundImage,5),",",",")
                    КонецЕсли;
                Если опСтр.className = "name" Тогда
                    новСтр.Наименование = опСтр.innerText;
                    новСтр.Ссылка = опСтр.firstChild.href;
                КонецЕсли;
                Если опСтр.className = "price" Тогда
                    новСтр.Цена = Число(СтрЗаменить(СтрЗаменить(опСтр.innerText,"-
",","),",",","));
                КонецЕсли;
                Если опСтр.className = "desc" Тогда
                    новСтр.Описание = опСтр.innerText;//опСтр.innerHTML - если
нужно вместе с тегами <br>
                КонецЕсли;
            КонецЦикла;

        КонецЕсли;

    КонецЦикла;
    Закреть();
КонецПроцедуры
```

Проверяю. Все работает.

Дело осталось за «операционной системой» и еще надо загрузить картинки. Давайте по порядку. Чтобы получить «ОС» нам надо открыть этот товар и считать «ОС» оттуда. Для этого делаем следующее:

- Добавляем на нижней панели кнопку «Доп Инфо» с кодом:

```
Для Каждого Стр из Товары Цикл
    Форма = ПолучитьФорму("ФормаДопИнфо");
    Форма.ЭлементыФормы.Сайт.Перейти(Стр.Ссылка);
    Форма.ТекущийТовар = Стр;
    Форма.ОткрытьМодально();
```

```
КонецЦикла;
```

- Создаем форму обработки «ФормаДопИнфо»
- Кладем на «ФормаДопИнфо» ПолеHTMLДокумента и называем его «Сайт»
- На событие ДокументСформирован у ПоляHTMLДокумента пишем код:

```
Если ЭлементыФормы.Сайт.Документ.body.all.length>1 Тогда
  ГрузимДопИнфо();
КонецЕсли;
```

- Создаем переменную в модуле формы

```
перем ТекущийТовар Экспорт;
```

- Создаем процедуру ГрузимДопИнфо():

```
Процедура ГрузимДопИнфо();
  Для Каждого Стр из ЭлементыФормы.Сайт.Документ.body.all Цикл
    Если Стр.className = "card_spec" Тогда
      Для Каждого опСтр из Стр.children Цикл
        Если Найти(опСтр.innerHTML, "Операционная система:") Тогда
          ТекущийТовар.ОС =
            СокрЛП(СтрЗаменить(опСтр.innerHTML, "Операционная система:", ""));
        КонецЕсли;
      КонецЦикла;
    КонецЕсли;
  КонецЦикла;
Закреть();
КонецПроцедуры
```

Проверяем и переходим к последнему пункту.

Заметили как похожи две последние инструкции? То-то же. Стремился к универсальности. Ну и наконец последний этап — Сохраним все изображения к примеру на диск «С» в папку «связной». Поехали

- Добавляем на нижней панели кнопку «Сохранить Картинки» с кодом:

```
Для Каждого Стр из Товары Цикл
  путьСамФайл = Лев(Стр.Картинка,Найти(Стр.Картинка, ".jpg/") + 3);
  самФайл =
  СтрЗаменить(СтрЗаменить(ПутьСамФайл, "http://static.svyaznoy.ru/upload/iblock/", ""), "/" , "");
  Стр.Картинка =
  СохранитьКартинкуСайта ("C:\Svyaznoy", ПутьСамФайл, СамФайл);
КонецЦикла;
```

- Добавляем функцию СохранитьКартинкуСайта:

```
Функция
```

```
СохранитьКартинкуСайта(КаталогСохранения,КартинкаНаСайте,КартинкаУНас)
ИмяФайлаКартинки = КаталогСохранения + "\" + КартинкаУНас;
ГетЗапрос = Новый СОМОбъект("WinHttp.WinHttpRequest.5.1");
ГетЗапрос.SetTimeouts(10000, 10000, 10000, 10000);
БазовыйУРЛ = КартинкаНаСайте;
Хидер1 = "Content-Type";
Хидер2 = "image/jpg"; // Тип рисунка.
ГетЗапрос.Open("GET", БазовыйУРЛ, False); // Синхронный режим.
ГетЗапрос.setRequestHeader(Хидер1, Хидер2);
ГетЗапрос.Send();
СтатусОтправки = ГетЗапрос.status;
Если СтатусОтправки <> 200 Тогда
    Сообщить("Ошибка отправки запроса на: "
        + КартинкаНаСайте);
    Возврат "";
КонецЕсли;

Стрим = Новый СОМОбъект("ADODB.Stream");
Стрим.Mode = 3;
Стрим.Type = 1;
Стрим.Open();
Стрим.Write(ГетЗапрос.responseBody);

Стрим.SaveToFile(ИмяФайлаКартинки, 2);
Стрим.Close();
Возврат ИмяФайлаКартинки;
КонецФункции
```

На этом наша эпопея с парсингом закончена. Это всего лишь пример того, как это можно сделать. Приложив сюда немного своего кода — можно сделать парсер для любого сайта.

[1С, 8.2, парсинг](#)

From:  
<https://wiki.rtzra.ru/> - RTzRa's hive

Permanent link:  
<https://wiki.rtzra.ru/software/1c/1c-parse-site>

Last update: **2017/05/09 18:34**

