



FlexiLayout Studio 8.0

Обучение на примерах
Пример 3

Содержание

Краткое описание документа Invoices.....	3
Шаг 1. Предварительные настройки.....	4
Шаг 2. Просмотр изображений и результатов предраспознавания	5
Шаг 3. Блоки.....	6
Шаг 4. Анализ изображений и определение последовательности поиска элементов и групп элементов	7
Шаг 5. Поиск заголовка поля Delivery Address, элемент kwDeliveryAddress	8
Шаг 6. Поиск заголовка поля Invoice Number, элемент kwInvoiceNumber	8
Шаг 7. Поиск заголовка поля Invoice Date, элемент kwInvoiceDate	10
Шаг 8. Описание поля Invoice Number, элемент InvoiceNumber	11
Шаг 9. Описание поля Invoice Date, элементы grDate, InvoiceDate и InvoiceDateAsString.....	12
Шаг 10. Описание составного элемента grAddress	15
Шаг 11. Поиск правой границы поля Delivery Address, элемент wgAddressRight	16
Шаг 12. Описание поля Delivery Address, элемент DeliveryAddress.....	17
Шаг 13. Продолжаем анализ изображений.....	18
Шаг 14. Поиск вспомогательного горизонтального сепаратора, элемент hsTableHeaderTop	19
Шаг 15. Анализ критериев поиска для заголовков столбцов таблицы. Составной элемент TableHeader	20
Шаг 16. Поиск заголовка столбца Quantity, элемент kwQuantity.....	21
Шаг 17. Поиск заголовка столбца UnitPrice, элемент kwUnitPrice.....	21
Шаг 18. Поиск заголовка столбца Total, элемент kwTotal	22
Шаг 19. Поиск заголовка столбца Sales, элемент kwSales.....	22
Шаг 20. Описание группы Footer, элемент Footer	22
Шаг 21. Описание итоговой части таблицы, элемент kwFooter	23
Шаг 22. Описание заголовка полей Total, элемент kwTotal	23
Шаг 23. Поиск заголовка поля Country, элемент kwOrigin	23
Шаг 24. Описание поля Country, элемент Country.....	24
Шаг 25. Поиск полей TotalQuantity и TotalAmount, элементы TotalQuantity и TotalAmount	25
Шаг 26. Поиск табличного элемента, элемент InvoiceTable	26
Шаг 27. Экспорт гибкого описания и подключение его к шаблону FlexiCapture.....	33

Краткое описание документа Invoices

Внимание! Для простоты в данном примере рассматривается одностраничный документ.

В данном руководстве ABBYY FlexiLayout Studio описывается процесс создания гибкого описания для простого счета. Её особенностью по сравнению с предыдущими примерами является наличие таблицы.

На этом примере мы проследим по шагам процесс создания гибкого описания, попутно знакомясь с некоторыми нестандартными подходами к описанию местоположения объектов, и, главное, научимся находить объекты изображения, соответствующие элементам типа Table.

Это руководство содержит ссылки на разделы справки Help/Tutorial/Sample1 и Help/Tutorial/Sample2, где описано создание гибких описаний с блоками и элементами различных типов (кроме табличных) и многие базовые операции более детализированы, а так же ссылки на другие разделы справки.

Изображения документов хранятся в

<имя диска>:\Documents and Settings\All Users\Application
Data\ABBYY\FlexiCapture\8.0\Samples\FlexiLayoutStudio\Invoice One-Page.

Reference	Description	Unit	Qty	Net unit price	Comment	Total CHF	Sales tax
AB.DD.GRAPHITE	Film 4C for the publication "Facing World" Visuel: PMS® 02 05 - Ref. 20040021 Graphic designer honoraries 860010	Unit	1	400.00		400.00	214
AB.DD.GRAPHITE	Film 4C for the publication "Talkies" Visuel: PMS® 01 03 - Ref. 20040033 Graphic designer honoraries 860010	Unit	1	400.00		400.00	214
AB.DD.GRAPHITE	Cost by DHL Graphic designer honoraries 860010	Unit	1	150.00		150.00	214
Total CHF						950.00	
Sales tax total							91.70
Total CHF						1041.70	
Sales tax summary							
Sales tax code	Rate	Total excluding tax	Sales tax	Total			
214	Uvsaorm	8.80	950.00	81.70	1031.70		
			950.00	81.70	1031.70		

Thanking you for your confidence, we would be more than willing to provide any further information you might require and remain, yours-most sincerely

J. Goldsmith manufacture
 ++41 235 23 13 - Fax ++41 235 22 13
 Internet: www.goldsmith-manufacture.com - E-mail: info@goldsmith-manufacture.com

Конечной целью создания гибкого описания данного документа в ABBYY FlexiLayout Studio является устойчивое нахождение на всех тестовых страницах местоположения следующих полей (в случае их присутствия на данном экземпляре документа):

Invoice Number
 Invoice Date
 Delivery Address
 Total Quantity
 Total Amount
 Country
 Invoice Table (только столбцы Reference, Designation, Quantity, Unit Price, Total)

J. Goldsmith manufacture

Delivery address: 70, Avenue Ampère, Montigny-le-Bretonneux, F-78886 Saint-Quentin, Yvelines Cedex, France

Billing address: 70, Avenue Ampère, Montigny-le-Bretonneux, F-78886 Saint-Quentin, Yvelines Cedex, France

Invoice no: 20040201

Date: 20040201

Customer: 253003

Sales tax: Export

Payment condition: 25 days net

Due date: 21.03.2004 CHF 765.20

Reference	Designation	Unit	Qty	Net amount	Discount	Unit Price	Total	Sales tax
AB.D0.GRAPHITE	Film 4C for the publication "Racing World" Visual P16P 02 05 - Ref. 20040201 Graphic designer honoraries 900010	Unit	1	400.00		400.00	400.00	214
AB.D0.GRAPHITE	Film 4C for the publication "Talkies" Visual P16P 02 05 - Ref. 20040201 Graphic designer honoraries 900010	Unit	1	400.00		400.00	400.00	214
AB.D0.GRAPHITE	Cost by GHI Graphic designer honoraries 900010	Unit	1	150.00		150.00	150.00	214
Total CHF								
Sales tax total							81.70	
Total CHF								1056.70

Sales tax summary

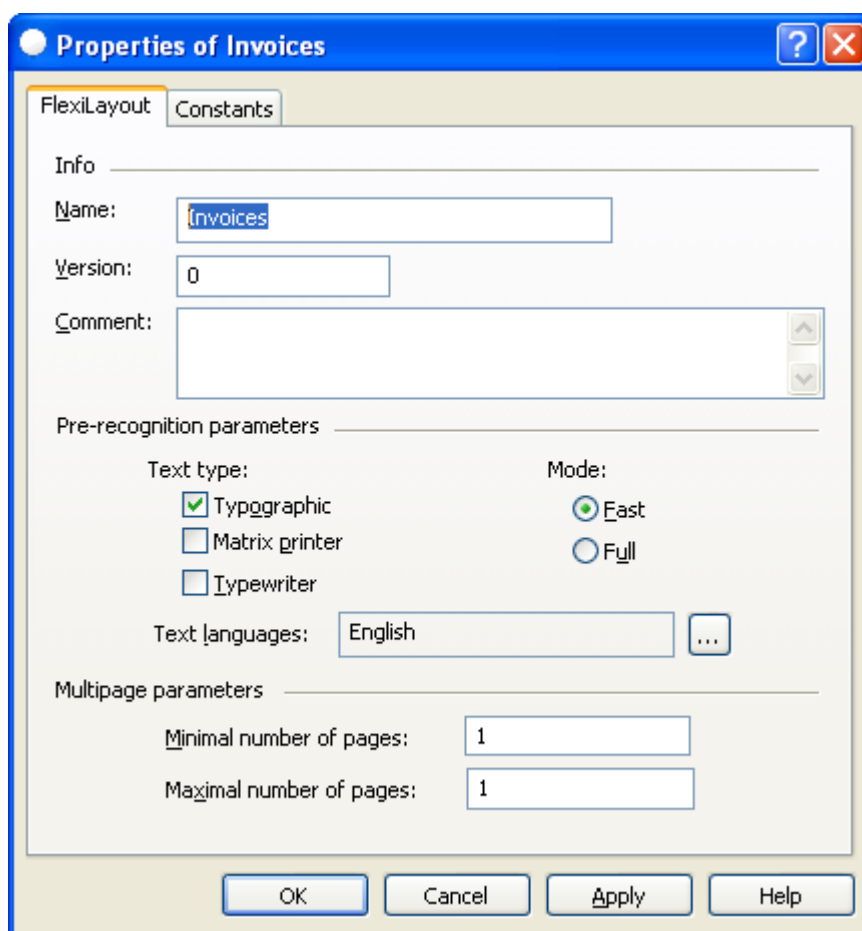
Sales tax code	Rate	Total excluding tax	Sales tax	Total
214	Livraison	850.00	81.70	1056.70
		850.00	81.70	1056.70

Thanking you for your coincidence, we would be more than willing to provide any further information you might require and remain, yours most sincerely

J. Goldsmith manufacture
 ++11 235 12 12 - Fax ++11 235 12 13
 Internet: www.goldsmithsm manufacture.com - E-mail: info@goldsmithsm manufacture.com

Шаг 1. Предварительные настройки

1. Создайте новую папку **Invoices One-Page**.
2. Запустите программу ABBYY FlexiLayout Studio 8.0.
3. Создайте новый проект и задайте имя проекта Sample3 (подробнее см. Пример 1, шаг 1).
4. Добавьте изображения в пакет (подробнее см. Пример 1, шаг 2).
Замечание: тестовые изображения третьего примера хранятся в папке <имя диска>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\8.0\Samples\FlexiLayoutStudio\Invoice One-Page\Invoice One-PageBatch.
5. Настройте свойства гибкого описания (подробнее см. Пример 1, шаг 3), учитывая, что тип печати – **Typographic**, а язык текста – **English**. Обратите внимание, что программа позволяет указать более одного языка. Задайте название гибкого описания **Invoice**. Режим предраспознавания оставьте **Fast**, предлагаемый по умолчанию, значения параметров **Minimal number of pages** и **Maximal number of pages** задайте равными 1.



Замечание. Более подробно о критериях выбора параметров предраспознавания см. Пример 1, шаг 3.

Для одностраничных документов нет необходимости в описании начала и конца документа с помощью определенных составных элементов **Header** и **Footer**. Поэтому их можно удалить из списка элементов гибкого описания.

Замечание. Для определения начала и конца документа в гибкое описание могут быть добавлены специальные элементы: **Header** и **Footer**. Если в процессе наложения гибкого описания не найден ни элемент **Header** ни элемент **Footer**, то число страниц в документе считается равным наибольшему числу страниц, указанному в свойствах гибкого описания. Для данного документа в свойствах гибкого описания мы указали 1.

Шаг 2. Просмотр изображений и результатов предраспознавания

Перед созданием элементов необходимо понять, на какие объекты изображений формы мы можем опереться для поиска ее полей. Опорными объектами могут служить только такие графические и/или текстовые фрагменты изображения, которые устойчиво находятся в процессе предраспознавания изображений формы.

Запустите предраспознавание изображений **Prerecognize** (подробнее см. Пример 1, шаг 4). Предварительный просмотр результатов предраспознавания всех тестовых страниц показывает, что, поскольку все текстовые объекты, соответствующие заголовкам искомым полем (для полей с заголовком), заголовки колонок таблицы и горизонтальные разделители обнаруживаются устойчиво, на них можно будет опираться при поиске других объектов изображения.

Замечание. Более подробно о просмотре страниц и результатах предраспознавания см. Пример 1, шаг 5, шаг 6. Теперь можно приступать непосредственно к созданию блоков и элементов. Для этого перейдем на окно **FlexiLayout** в главном окне программы ABBYY FlexiLayout Studio.

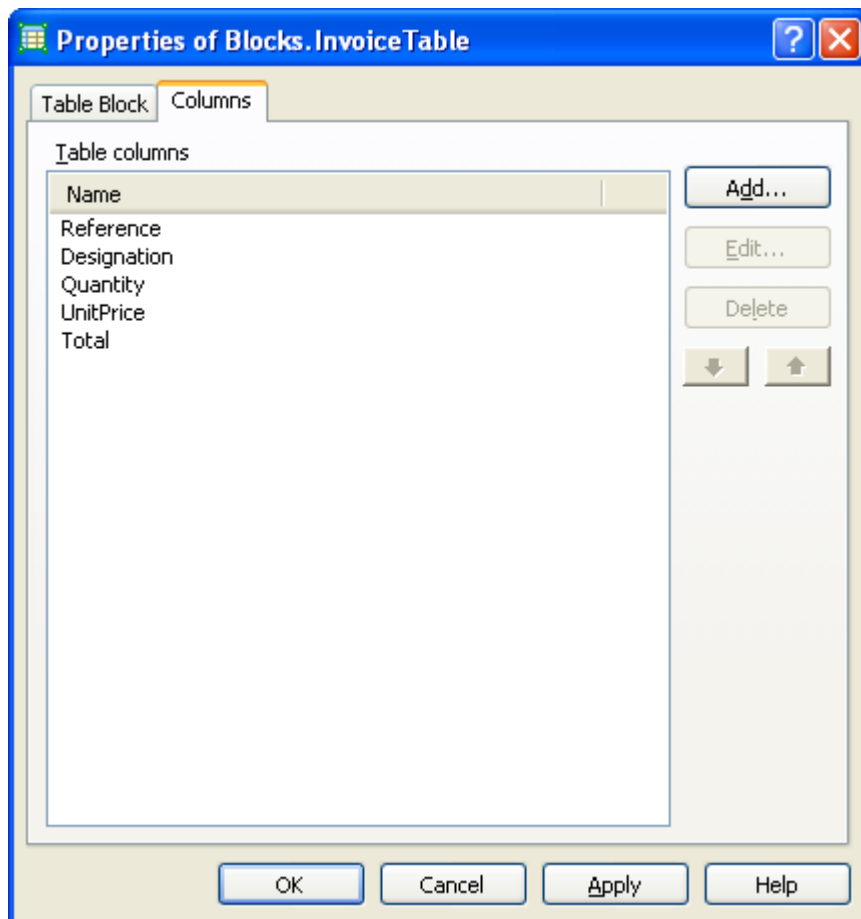
Шаг 3. БЛОКИ

Целью создания данного гибкого описания является нахождение местоположения следующих блоков (в случае их наличия на изображении):

- **Invoice Number**
- **Invoice Date**
- **Delivery Address**
- **Total Quantity**
- **Total Amount**
- **Country**
- **Invoice Table** (только столбцы **Reference, Designation, Quantity, Unit Price, Total**)

Поэтому начнем с создания этих блоков, чтобы в процессе создания гибкого описания иметь их список перед глазами. Для этого:

1. Создайте блок **InvoiceNumber** (тип **Text**), соответствующий полю **Invoice Number**.
2. Создайте блок **InvoiceDate** (тип **Text**), соответствующий полю **Invoice Date**.
3. Создайте блок **DeliveryAddress** (тип **Text**), соответствующий полю **Delivery Address**.
4. Создайте блок **TotalQuantity** (тип **Text**), соответствующий полю **Total Quantity**.
5. Создайте блок **TotalAmount** (тип **Text**), соответствующий полю **Total Amount**.
6. Создайте блок **Country** (тип **Text**), соответствующий полю **Country**.
7. Создайте блок **InvoiceTable** (тип **Table**), соответствующий полю **Invoice Table**. В диалоге свойств блока на закладке **Columns** следует задать такие названия столбцов и такой порядок их следования, какой мы хотим видеть позже у данного блока в шаблоне FlexiCapture. Последовательно добавьте названия столбцов в следующем порядке: **Reference, Designation, Quantity, UnitPrice, Total** (подробнее см. Справка\Работа с таблицами\Блок типа Table).



Замечание. На данном этапе мы указываем только тип блока и название (и список столбцов для блока типа **Table**). Местоположение каждого блока мы пропишем позже, когда все вспомогательные элементы будут описаны и отлажены. Важно помнить, что *блок* типа **Table** обязательно должен быть создан раньше, чем *элемент* типа **Table**, с помощью которого мы планируем искать этот блок, поскольку элемент типа **Table** должен иметь ссылку на блок, для поиска которого он предназначен. Этим он отличается от блоков типа **Text**, **Barcode**, **Checkmark** или **Picture**, которые можно создавать в процессе разработки гибкого описания, после того, как все элементы для описания местоположения того или иного блока будут созданы и отлажены (см. Пример 1, Пример 2).

На этом процедура задания свойств гибкого описания и начальных приготовлений к работе закончена и можно приступать непосредственно к созданию элементов.

Шаг 4. Анализ изображений и определение последовательности поиска элементов и групп элементов

Нам нужно решить:

- Какова закономерность расположения полей на изображениях?
- На какие элементы мы можем опереться в поиске полей?
- Какую выбрать последовательность поиска (поскольку на каждом следующем шаге мы можем опираться только на описанные выше элементы)?

Проанализируем наши изображения.

Мы видим, что центральная часть документа отдана под таблицу, которая присутствует на всех изображениях.

Замечание. Напомним, что в терминологии FlexiLayout Studio 8.0 таблицей является объект изображения, состоящий из фрагментов, расположенных в виде колонок и строк, визуально разделённых сепараторами или просветами.

Выше таблицы располагается группа полей (назовём её **Header**) **Invoice Number**, **Invoice Date** и **Delivery Address**. При этом поле **Invoice Number** присутствует всегда и его можно сделать обязательным (идентификационным) полем, в то время как поля **Invoice Date** и **Delivery Address** могут и отсутствовать. Под таблицей располагается группа полей (назовём её **Footer**) **TotalQuantity**, **TotalAmount** и **Country**. Эти поля тоже присутствуют не всегда.

Начнём строить наш поиск с верхней части документа. Для поиска полей, расположенных в верхней части, создадим логическую группу, в которую войдут все элементы для поиска полей **Invoice Number**, **Invoice Date** и **Delivery Address**.

1. Создайте составной элемент **InvoiceHeader** (тип **Group**).

Мы видим, что поля **Invoice Number**, **Invoice Date** и **Delivery Address** всегда располагаются в левом верхнем углу, при этом порядок их следования постояен: сначала идёт **Delivery Address**, потом **Invoice Number**, потом **Invoice Date** (конечно, при условии, что они присутствуют на изображении). В таком порядке и будем их искать. Для описания поиска ключевых слов (**keywords**) заголовков полей **Invoice Number**, **Invoice Date** и **Delivery Address** воспользуемся элементами типа **Static Text**. Для этого в составе **InvoiceHeader** нам необходимы элементы:

2. элемент **kwDeliveryAddress** (тип **Static Text**), соответствующий заголовку поля **Delivery Address** (подробная инструкция в Шаге 5).
3. элемент **kwInvoiceNumber** (тип **Static Text**) соответствующий заголовку поля **Invoice Number** (подробная инструкция в Шаге 6).
4. элемент **kwInvoiceDate** (тип **Static Text**), соответствующий заголовку поля **Invoice Date** (подробная инструкция в Шаге 7).
Что касается поиска самих полей **Invoice Number** и **Invoice Date**, будем искать их в той же строке, что и соответствующий заголовок, правее заголовка. В составе элемента **InvoiceHeader** нам потребуются элементы:
5. элемент **InvoiceNumber** (тип **Character String**), соответствующий полю **Invoice Number** (подробная инструкция в Шаге 8).
6. составной элемент **grDate** (тип **Group**) для поиска поля **Invoice Date** (подробная инструкция в Шаге 9). В составе элемента **InvoiceHeader**, **grDate** будем описывать элементы:
7. элемент **InvoiceDate** (тип **Date**), соответствующий полю **Invoice Date** в случае хорошего качества изображения (подробная инструкция в Шаге 9).
8. элемент **InvoiceDateAsString** (тип **Character String**), соответствующий полю **Invoice Date** в том случае, если поиск элемента **InvoiceDate** оказался неуспешным (подробная инструкция в Шаге 9).
9. Поле **Delivery Address** многострочное, поэтому нам потребуется элемент типа **Text Fragment**, но прежде нам необходимо будет задать область поиска самым строгим образом. В качестве ограничителя области

поиска справа будем использовать элемент типа **White Gap**. И объединим все эти элементы, служащие описанию местоположения поля **Delivery Address**, в отдельную группу:

В составе элемента **InvoiceHeader** создадим составной элемент **grAddress** (тип **Group**) (подробная инструкция в Шаге 10).

В составе элемента **grAddress** опишем элементы:

10. вспомогательный элемент **wgAddressRight** (тип **White Gap**), прилегающий к полю **Delivery Address** справа (подробная инструкция в Шаге 11).
11. элемент **DeliveryAddress** (тип **Text Fragment**), соответствующий полю **Delivery Address** (подробная инструкция в Шаге 12).

Шаг 5. Поиск заголовка поля Delivery Address, элемент kwDeliveryAddress

Проанализировав изображения, видим, что ключевыми словами для поиска заголовка поля **Delivery Address** будет однострочная строка «DeliveryAddress». Число ошибок в заголовке ограничим 20 процентами, что в нашем случае будет означать не более 3х ошибок во всей фразе.

Замечание. Регистр написания не влияет на поиск и используется только для повышения читабельности. Допустимое число ошибок подбирается опытным путем. Более подробно о параметрах элемента типа Static Text см. Справка Элементы/Свойства элементов, задающие параметры искомого объекта/Элемент Static Text.

Для того чтобы создать элемент **kwDeliveryAddress**:

1. В составе элемента **InvoiceHeader** создайте элемент **kwDeliveryAddress** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **DeliveryAddress**.
4. Задайте значение свойства **Max error percentage** равным 20.
5. Выберите закладку **Advanced**.
6. Поскольку поле **Delivery Address** находится в верхней трети изображения по вертикали и в левой трети по горизонтали, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения в левой трети страницы. Искать объект изображения в верхней трети страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:

Above: PageRect.Top + PageRect.Height/3;
LeftOf: PageRect.Left + PageRect.Width/3;
7. Протестируйте изображения и убедитесь, что элемент найден на всех изображениях, где он присутствует.

Замечание. Чтобы наложить гибкий шаблон на все тестовые изображения, удобно воспользоваться горячими клавишами.

Для этого:

- Поместите фокус в окно пакета изображений **Batch**.
- Нажмите клавиши **Ctrl + A** для выбора всех страниц пакета.
- Нажмите клавиши **Ctrl + Shift + E** для запуска анализа выделенных страниц.

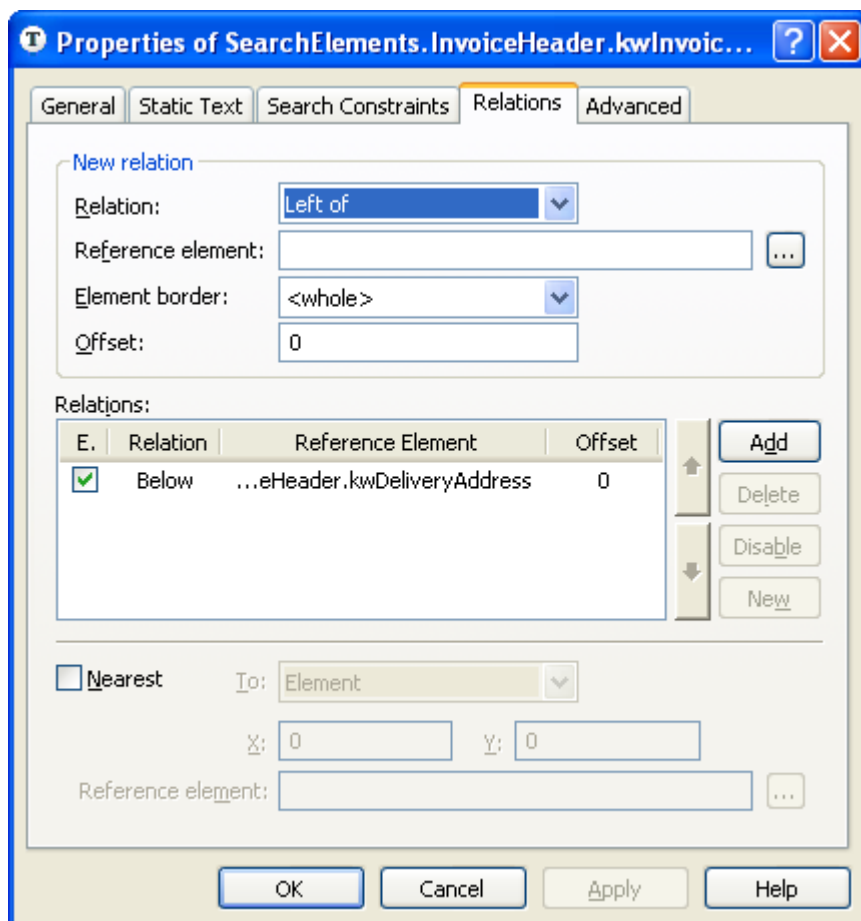
Шаг 6. Поиск заголовка поля Invoice Number, элемент kwInvoiceNumber

Проанализировав изображения, видим, что заголовок поля **Invoice Number** присутствует на всех изображениях и отличается от других типов документов. При обработке гибких форм в программе ABBYY FlexiCapture хотелось бы по возможности полностью исключить случайное распознавание форм, не относящихся к заданному типу. Одним из способов идентификации формы является задание, по крайней мере, одного обязательного элемента. Поэтому сделаем данный элемент обязательным, т.е. идентификационным элементом для данного типа форм.

Так же заметим, что заголовок на части изображений представлен фразой **Invoice no.**, а на части – **Credit note no.** Составим нашу поисковую строку из 2х вариантов, при этом запишем каждую фразу слитно, чтоб ограничения на количество ошибок относились ко всей фразе целиком, а не к отдельным маленьким словам.

Для того чтобы создать элемент **kwInvoiceNumber**:

1. В составе элемента **InvoiceHeader** создайте элемент **kwInvoiceNumber** (тип **Static Text**).
2. Отмените заданное по умолчанию свойство **Optional** на закладке **General** (поскольку данный элемент мы решили сделать обязательным).
3. Выберите закладку **Static Text**.
4. В поле **Search Text** укажите значение искомого текста: **InvoiceNo|CreditNoteNo**
5. Задайте значение свойства **Max error percentage** равным 20.
6. Выберите закладку **Relations**.
7. На закладке **Relations** зададим, что ключевые слова заголовка поля **Invoice Number** могут находиться только ниже заголовка поля **Delivery Address**.
Below элемента **InvoiceHeader.kwDeliveryAddress**, **Offset = 0**



8. Выберите закладку **Advanced**.
9. Поскольку поле **Invoice Number** находится в верхней трети изображения по вертикали и в левой трети по горизонтали, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения в левой трети страницы. Искать объект изображения в верхней трети страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:

Above: PageRect.Top + PageRect.Height/3;
LeftOf: PageRect.Left + PageRect.Width/3;

10. Протестируйте изображения и убедитесь, что элемент найден на всех изображениях.

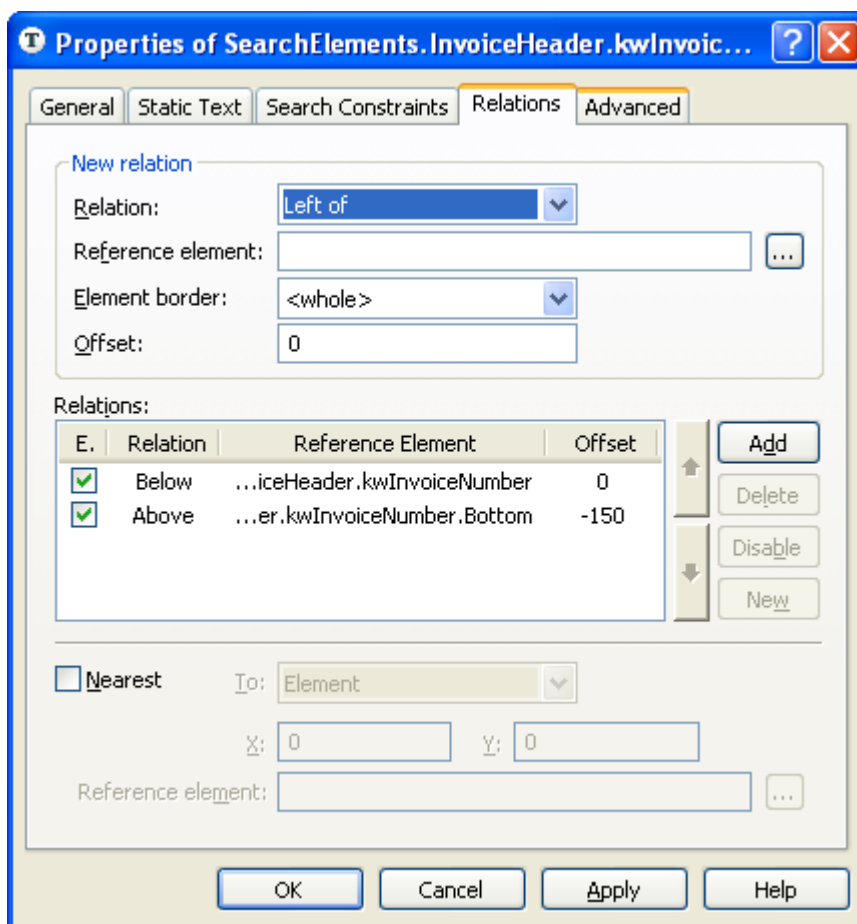
Замечание. В реальной практике подобная ситуация, когда текст заголовка может быть представлен более чем одним вариантом, является сигналом: нужно убедиться, что учтены все варианты написания заголовка, или хотя бы подавляющее большинство их. Можно запросить больше тестовых изображений или постараться узнать у клиента каким-то другим способом, какие варианты написания заголовка поля могут присутствовать на данном типе форм.

Шаг 7. Поиск заголовка поля Invoice Date, элемент kwInvoiceDate

Проанализировав тестовые изображения, видим, что заголовок поля Invoice Date представлен коротким и очень распространенным словом **Date**. Это означает, что нам нужно постараться максимально сузить область поиска. Пролитав изображения, замечаем, что заголовок поля **Date** (в случае наличия на изображении этого поля) располагается очень близко к заголовку поля **Invoice Number**.

Для того чтобы создать элемент **kwInvoiceDate**:

1. В составе элемента **InvoiceHeader** создайте элемент **kwInvoiceDate** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **Date**
4. Оставьте значение свойства **Max error percentage** равным 30 (что для нашего 4х буквенного слова означает ошибку не более чем в одной букве).
5. Выберите закладку **Relations**.
6. На закладке **Relations** зададим, что ключевые слова заголовка могут находиться только ниже заголовка поля **Invoice Number**, но не ниже чем на расстоянии 150 dot от его нижней границы (**Below** элемента **kwInvoiceNumber**, **Offset** = 0 и **Above** нижней границы региона элемента **kwInvoiceNumber**, **Offset** = -150, **Element border** = Bottom).



Замечание. При подборе значений для отступов (offset) бывает полезно обратить внимание на геометрические свойства (границы и размеры) опорных и искомых объектов изображения в режиме просмотра результатов предраспознавания. Подбор значений для отступов (offset) является творческим процессом и осуществляется путем проб и ошибок.

7. Выберите закладку **Advanced**.
8. Поскольку поле **Invoice Date** находится в левой половине по горизонтали, в поле **Advanced pre-search relations** задайте ещё одно дополнительное условие поиска: *Искать объект изображения в левой половине страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:



LeftOf: PageRect.Left + PageRect.Width/2;

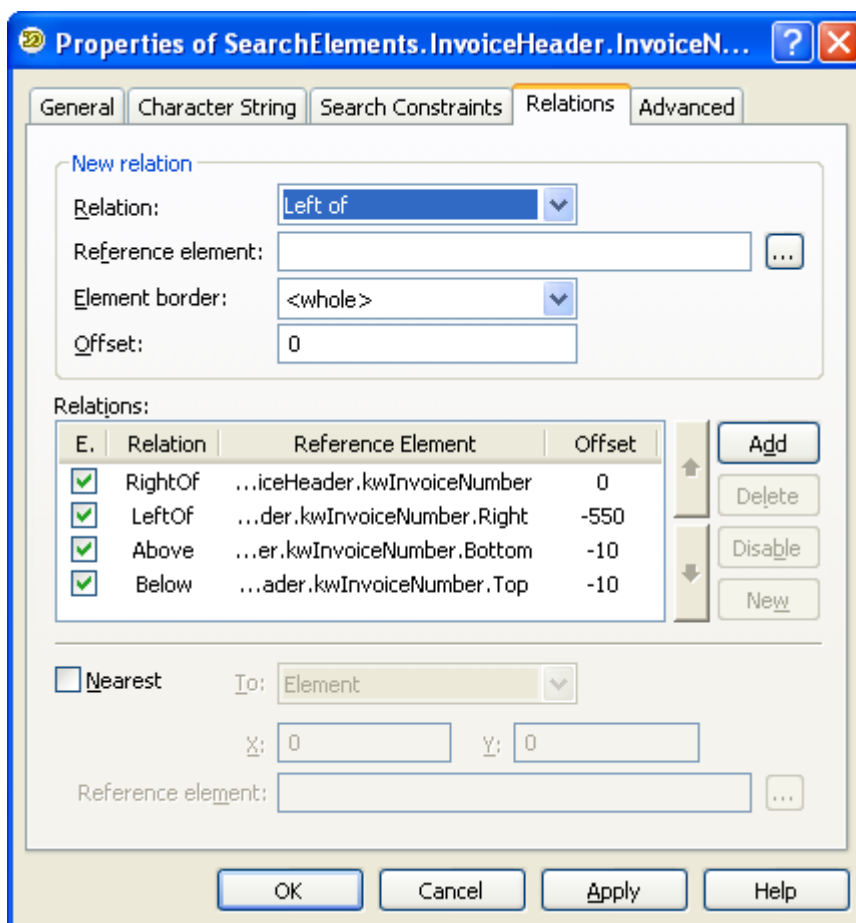
9. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях, где он присутствует.

Шаг 8. Описание поля Invoice Number, элемент InvoiceNumber

Поле **Invoice Number** располагается правее заголовка и на том же уровне. Проанализировав наши тестовые изображения, видим, что на некоторых текст номера распознаётся очень плохо. Поэтому не будем накладывать ограничений на алфавит, но ограничим область поиска и длину цепочки 3–20 символами.

Для того чтобы создать элемент **InvoiceNumber**:

1. В составе элемента **InvoiceHeader** создайте элемент **InvoiceNumber** (тип **Character String**).
 **Замечание.** Совпадение имени элемента с именем блока, соответствующего полю **InvoiceNumber**, не обязательно, но удобно для дальнейшей работы с гибким описанием.
2. Выберите закладку **Character String**.
3. Задайте значение параметра **Percentage of non-alphabet characters** равным 100%, что означает отсутствие ограничений на входящие символы.
4. В поле **String Length** задайте нечеткий интервал { -1, 3, 20, INF }, оценивающий длину цепочки символов. Предполагаем, что все возможные значения будут находиться в интервале от 3 до 20 символов. Любая гипотеза, выходящая за рамки данного интервала, будет штрафоваться.
5. Перейдите на закладку **Relations**.
6. Поскольку **InvoiceNumber** может находиться только справа от ключевых слов заголовка **kwInvoiceNumber**, но не правее чем 550 dot от правой границы заголовка **kwInvoiceNumber**, задайте следующее условие поиска:
 - **RightOf** элемента **kwInvoiceNumber**, **Offset** = 0
 - **LeftOf** элемента **kwInvoiceNumber**, **Offset** = -550, **Element border** = Right **Замечание.** Значения для отступов (offset) подбираются опытным путем.
7. Поскольку **InvoiceNumber** всегда находится на одном уровне с ключевыми словами заголовка **kwInvoiceNumber**, добавьте следующие условия поиска:
 - **Above** элемента **kwInvoiceNumber**, **Offset** = -10, **Element border** = Bottom
 - **Below** элемента **kwInvoiceNumber**, **Offset** = -10, **Element border** = Top



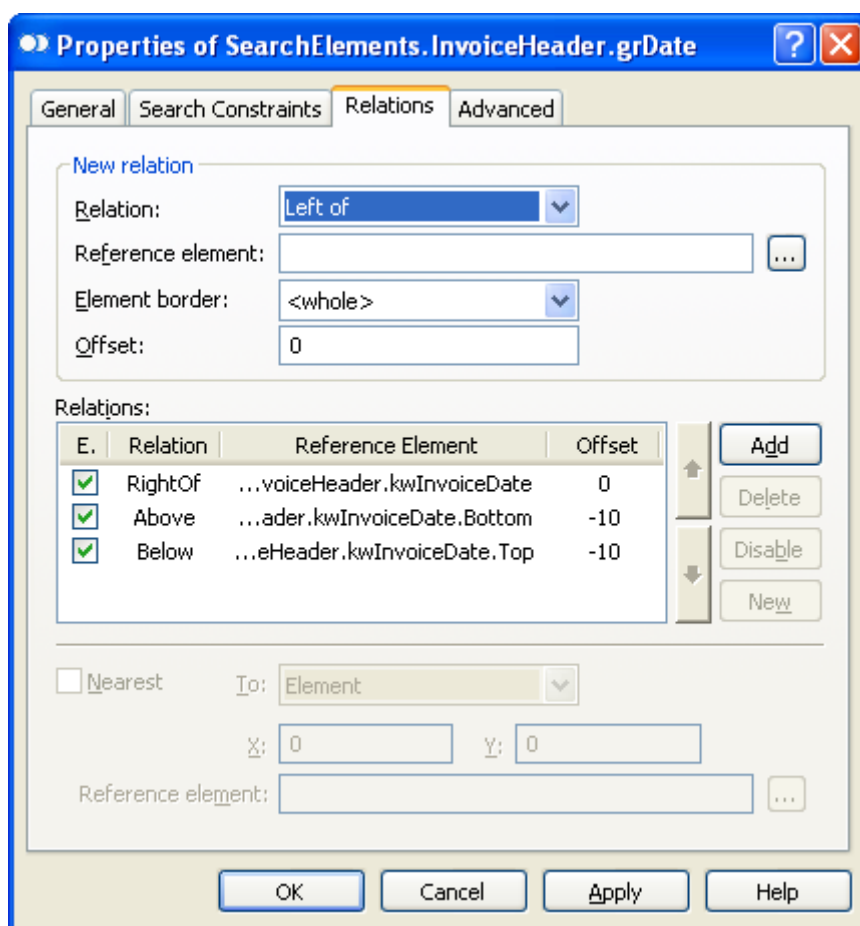
8. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях, где он присутствует.
9. Выберите в качестве описания местоположения блока **InvoiceNumber** опцию **Source element**. В качестве source element укажите элемент **InvoiceNumber**, нажав кнопку "...".

Шаг 9. Описание поля Invoice Date, элементы grDate, InvoiceDate и InvoiceDateAsString

Проанализировав изображения, видим, что поле **Invoice Date** располагается правее заголовка и на одном с ним уровне. Для поиска даты воспользуемся элементом типа **Date**. На случай появления изображений низкого качества, где текст даты может распознаваться не очень хорошо и поиск элемента типа **Date** не даст результатов, дополнительно опишем элемент типа **Character String**. Для того чтобы задать настройки, общие для всех этих элементов, нам потребуется составной элемент **grDate**:

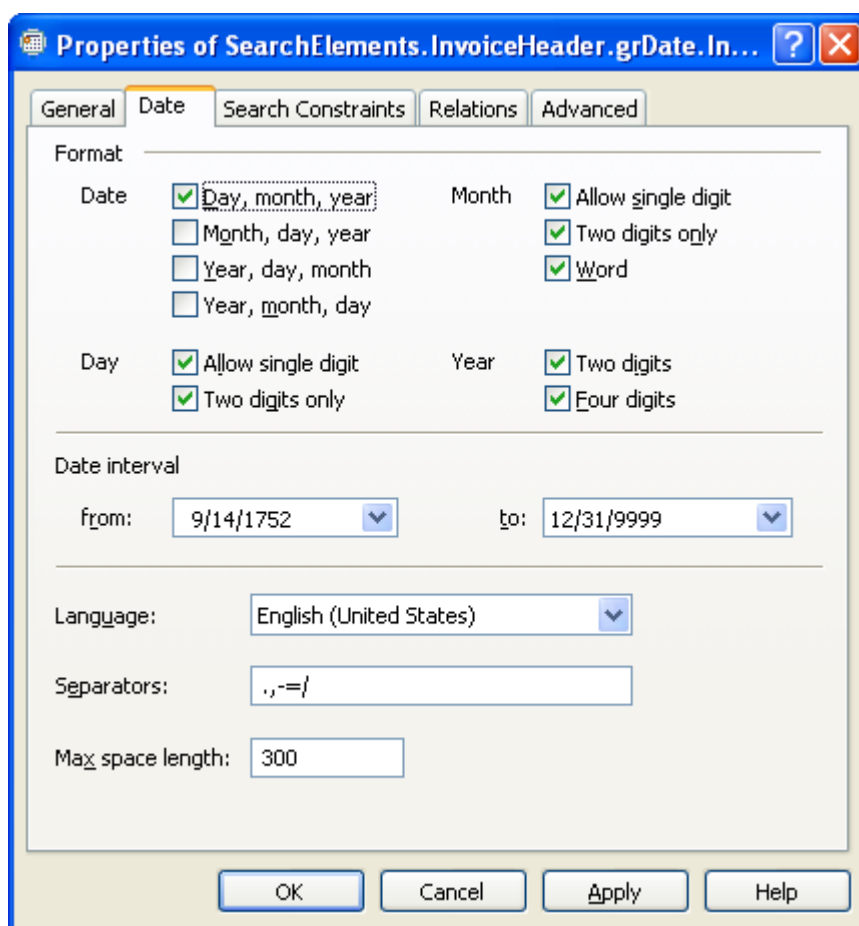
1. В составе элемента **InvoiceHeader** создайте элемент **grDate** (тип **Group**).
2. Перейдите на закладку **Relations**.
3. Поскольку все элементы, служащие поиску местоположения поля **Invoice Date**, могут находиться только правее ключевых слов заголовка **kwInvoiceDate** и на одном с ними уровне, задайте следующие условия поиска:

- **RightOf** элемента **kwInvoiceDate**, **Offset** = 0
- **Above** элемента **kwInvoiceDate**, **Offset** = -10, **Element border** = Bottom
- **Below** элемента **kwInvoiceDate**, **Offset** = -10, **Element border** = Top



Для того чтобы создать элемент **InvoiceDate**:

1. В составе элемента **InvoiceHeader.grDate** создайте элемент **InvoiceDate** (тип **Date**).
2. Выберите закладку **Date**.
3. Для элемента **InvoiceDate** на закладке **Date** следует указать все допустимые форматы записи даты. Предполагаем, что формат даты **Day, Month, Year**, язык **English**, других ограничений на дату нет.



4. Выберите закладку **Advanced**.
5. Поскольку поле **Invoice Date** может находиться на изображении, а может и нет, но при этом всегда присутствует вместе с заголовком и расположено ближе всего к заголовку в области поиска, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения, только если найден элемент **kwInvoiceDate**. Искать объект изображения, ближайший к элементу **kwInvoiceDate***. На языке описания FlexiLayout данное условие выглядит следующим образом:

```
If InvoiceHeader.kwInvoiceDate.IsNull Then DontFind;
Nearest: InvoiceHeader.kwInvoiceDate;
```

Замечание. Условие **Nearest** также можно задать через интерфейс программы. Для этого перейдите на закладку **Relations**, отметьте опцию **Nearest**, в выпадающем списке **To**: выберите **Element** и укажите **kwInvoiceDate** в качестве **Reference element**.

6. Протестируйте изображения и изучите, как верно программа находит элемент на всех изображениях, где поле Invoice Date присутствует.

Для того чтобы создать элемент **InvoiceDateAsString**:


1. В составе элемента **InvoiceHeader.grDate** создайте элемент **InvoiceDateAsString** типа **Character String**.
2. Выберите закладку **Character String**.
3. В поле **String Length** задайте нечеткий интервал { -1, 8, 14, INF }, оценивающий длину цепочки символов. Предполагаем, что все возможные значения будут находиться в интервале от 8 до 14 символов. Любая гипотеза, выходящая за рамки данного интервала, будет штрафоваться.
4. В поле **Max space length** задайте значение 20, ограничивающее допустимый размер просвета в текстовой строке в 20 dot;
5. Не накладывайте других ограничений на параметры элемента, оставьте значения по умолчанию.
6. Выберите закладку **Advanced**.
7. Поскольку данный элемент мы хотим искать только в том случае, если элемент **InvoiceDate** не обнаружен на изображении, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения, только если не найден элемент **InvoiceDate***. На языке описания FlexiLayout данное условие выглядит следующим образом:

If Not InvoiceDate.IsNotNull Then DontFind;


8. Поскольку поле **Invoice Date** может находиться на изображении, а может и нет, но при этом всегда присутствует вместе с заголовком и расположено ближе всего к заголовку в области поиска, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: Искать объект изображения, только если найден элемент **kwInvoiceDate**. *Искать объект изображения, ближайший к элементу **kwInvoiceDate***. На языке описания FlexiLayout данное условие выглядит следующим образом:

If InvoiceHeader.kwInvoiceDate.IsNotNull Then DontFind;

Nearest: InvoiceHeader.kwInvoiceDate;

 **Замечание.** Условие **Nearest** также можно задать через интерфейс программы. Для этого перейдите на закладку **Relations**, отметьте опцию **Nearest**, в выпадающем списке **To**: выберите **Element** и укажите **kwInvoiceDate** в качестве **Reference element**.

9. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях, где поле **Invoice Date** присутствует, и для элемента **InvoiceDate** сформирована гипотеза.
10. Выберите в качестве описания местоположения блока **InvoiceDate** опцию **Source element**. В качестве **source element** укажите элемент **InvoiceHeader.grDate**, нажав кнопку "..."

 **Замечание.** Регион составного элемента состоит из объединения регионов найденных подэлементов.

Шаг 10. Описание составного элемента grAddress

Местоположение поля **Delivery Address** мы решили искать с помощью следующих элементов: просвета – ограничителя справа и элемента типа текстовый фрагмент.

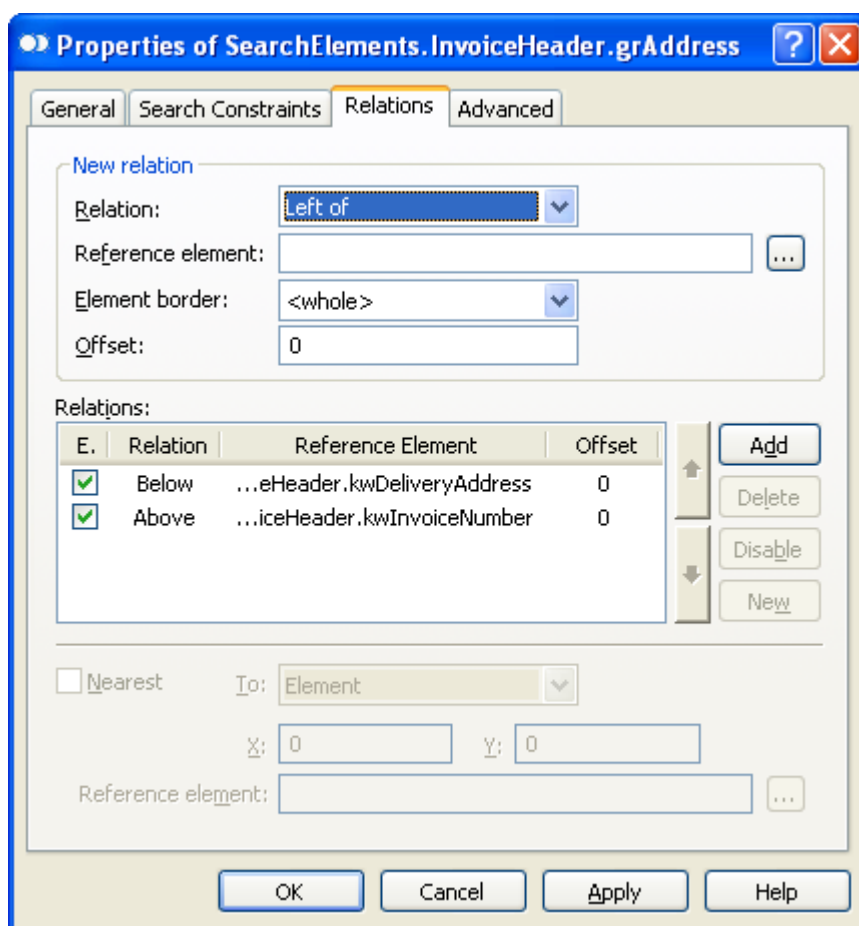
Для того чтобы задать настройки, общие для этих элементов, нам потребуется составной элемент.

Для того чтобы создать элемент **grAddress**:

1. В составе элемента **InvoiceHeader** создайте элемент **grAddress** (тип **Group**).
2. Перейдите на закладку **Relations**.
3. Поскольку все элементы, предназначенные для поиска местоположения поля **Delivery Address**, могут находиться только ниже от ключевых слов заголовка **kwDeliveryAddress**, и только выше ключевых слов заголовка **kwInvoiceNumber**, задайте следующие условия поиска:

– **Below** элемента **kwDeliveryAddress**, **Offset = 0**

– **Above** элемента **kwInvoiceNumber**, **Offset = 0**



4. Выберите закладку **Advanced**.
5. Поскольку поле **Delivery Address**, а так же ограничивающий его просвет, находятся в левых двух пятых страницы по горизонтали, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объекты изображения в левых двух пятых страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:

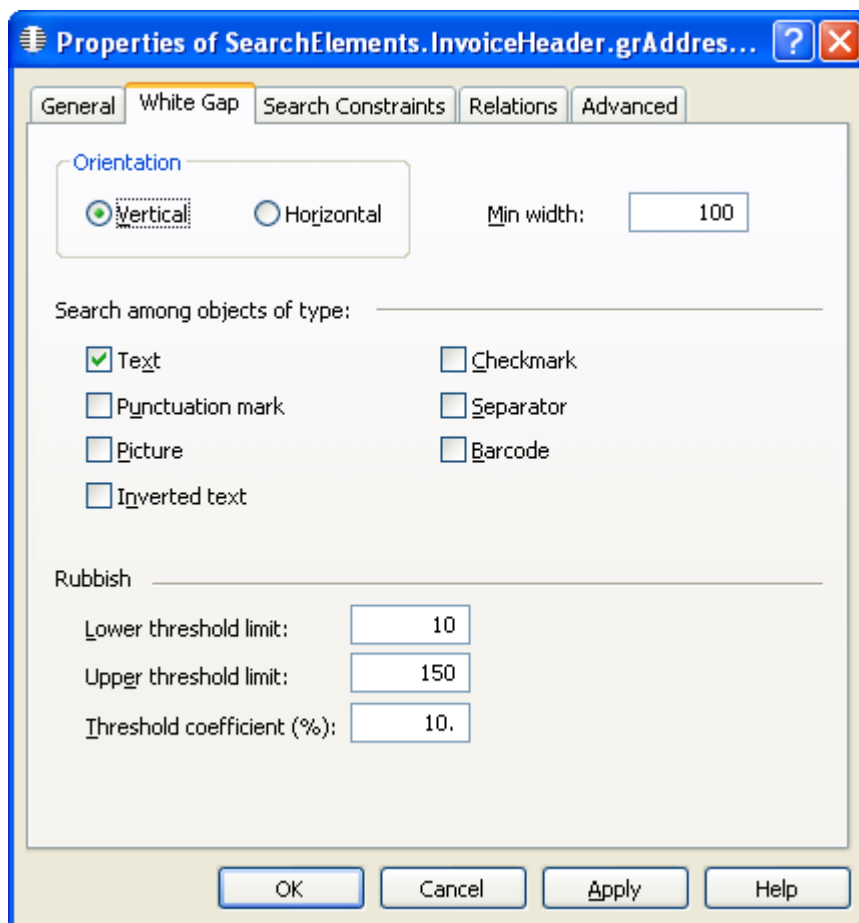
LeftOf: PageRect.Left + PageRect.Width*2/5;

Шаг 11. Поиск правой границы поля Delivery Address, элемент wgAddressRight

Для ограничения области поиска элемента **DeliveryAddress** справа опишем вертикальный просвет (white gap).

Для того чтобы создать элемент **wgAddressRight**:


1. В составе элемента **InvoiceHeader.grAddress** создайте элемент **wgAddressRight** (тип **White Gap**).
2. Выберите закладку **White Gap**.
3. В группе **Orientation** выберите **Vertical**. Минимальную ширину (**Min Width**) поставьте равной 100.
Замечание. При подборе значений для высоты/ширины просветов бывает полезно обратить внимание на геометрические свойства (границы и размеры) близлежащих объектов изображения в режиме просмотра результатов предраспознавания.
4. Параметры **Lower threshold limit**: и **Threshold coefficient (%)**: установите равными 10.
Замечание. Значения свойств **Lower threshold limit**, **Upper threshold limit** и **Threshold coefficient (%)** подбираются опытным путем.
 Подробнее см. Справка раздел Элементы/Свойства элементов, задающие параметры искомого объекта изображения/элемент White Gap
5. Не накладывайте других ограничений на параметры элемента, оставьте значения по умолчанию.



6. Перейдите на закладку **Relations**.
7. Поскольку просвет может находиться только справа от ключевых слов заголовка **kwDeliveryAddress**, задайте следующее условие поиска:
 - **RightOf** элемента **kwDeliveryAddress**, **Offset = 0**
8. Выберите закладку **Advanced**.
9. Поскольку поле **Delivery Address** может присутствовать на изображении, а может и нет, но при этом всегда присутствует вместе с заголовком, в поле **Advanced pre-search relations** задайте ещё одно дополнительное условие поиска: *Искать объект изображения, только если найден элемент **kwDeliveryAddress***. На языке описания FlexiLayout данное условие выглядит следующим образом:


```
If InvoiceHeader.kwDeliveryAddress.IsNull Then DontFind;
```
10. Поскольку интересующий нас просвет должен также прилегать к полю **Invoice Number**, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения ближайший к элементу **kwInvoiceNumber***. На языке описания FlexiLayout данное условие выглядит следующим образом:


```
Nearest: InvoiceHeader.kwInvoiceNumber;
```

 **Замечание.** Условие **Nearest** также можно задать через интерфейс программы. Для этого перейдите на закладку **Relations**, отметьте опцию **Nearest**, в выпадающем списке **To**: выберите **Element** и укажите **kwInvoiceNumber** в качестве **Reference element**.
11. Протестируйте изображения и убедитесь, что программа верно находит элемент **wgAddressRight** на всех изображениях, где присутствует поле **Delivery Address**.


Шаг 12. Описание поля Delivery Address, элемент DeliveryAddress

Для того чтобы создать элемент **DeliveryAddress**:

1. В составе элемента **InvoiceHeader.grAddress** создайте элемент **DeliveryAddress** (тип **Text Fragment**).
2. Перейдите на закладку **Relations**.

Поскольку **DeliveryAddress** может находиться только выше заголовка **kwInvoiceNumber**, только левее просвета **wgAddressRight**, только правее начала заголовка **kwDeliveryAddress** с небольшим запасом, задайте следующее условие поиска:

- **Above** элемента **kwInvoiceNumber**, **Offset = 0**
- **LeftOf** элемента **wgAddressRight**, **Offset = 0**
- **RightOf** элемента **kwDeliveryAddress**, **Offset = -50**, **Element border = Left**

 **Замечание.** При подборе значений для отступов (offset) бывает полезно обратить внимание на геометрические свойства (границы и размеры) опорных и искомых объектов изображения в режиме просмотра результатов предраспознавания. Значения для отступов (offset) подбираются опытным путем.

3. Поскольку условие, что поле **DeliveryAddress** может находиться только ниже от ключевых слов заголовка **kwDeliveryAddress**, уже задано на уровне группы **InvoiceHeader.grAddress**, мы не будем его дублировать здесь.
4. Выберите закладку **Advanced**.
5. Поскольку поле **Delivery Address** может присутствовать на изображении, а может и нет, но при том всегда присутствует вместе с заголовком, в поле **Advanced pre-search relations** задайте ещё одно дополнительное условие поиска: *Искать объект изображения, только если найден элемент **kwDeliveryAddress***. На языке описания FlexiLayout данное условие выглядит следующим образом:

```
If InvoiceHeader.kwDeliveryAddress.IsNull Then DontFind;
```

6. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях, где он присутствует.
7. Выберите в качестве описания местоположения блока **DeliveryAddress** опцию **Source element**. В качестве source element укажите элемент **DeliveryAddress**, нажав кнопку "...".

Шаг 13. Продолжаем анализ изображений

К этому моменту у нас описаны поля **Invoice Number**, **Invoice Date** и **DeliveryAddress**.

Продумаем последовательность дальнейшего поиска элементов.

Изучая структуру расположения полей на тестовых изображениях дальше, замечаем, что для поиска поля **Country** нам, скорее всего, будет достаточно заголовка этого поля и элемента типа **Character String. B** то же время для поиска полей **Total Quantity** и **Total Amount** одного заголовка, общего для этих полей, явно недостаточно, дополнительно потребуются опорные элементы для локализации полей по горизонтали. В качестве таких элементов могут служить заголовки соответствующих столбцов таблицы.

Что касается поля **Invoice Table**, которое мы будем искать с помощью элемента типа **Table**, здесь нам тоже понадобятся предварительные приготовления.

Во-первых, Header таблицы. Определить местоположение заголовков столбцов можно двумя способами. Первый способ – указать одно или несколько ключевых слов, открыв диалог свойств столбца с закладки **Columns** в диалоге свойств табличного элемента. Этот способ более быстрый в смысле разработки, но менее гибкий. Второй способ – использовать ранее найденный элемент для определения местоположения заголовка. Здесь в нашем распоряжении вся гибкость настроек, если она понадобится. Проанализировав наши изображения, видим, что качество хорошее и нет никаких особенностей, которые могут помешать выбрать первый метод (разве что очень короткий заголовок столбца **Quantity**: для 3х буквенного ключевого слова допустимый процент ошибок 20% означает, что ни одной ошибки не допускается). Решаем для заголовков первых 2х столбцов **Reference** и **Designation** воспользоваться первым способом, а для остальных – столбцы **Quantity**, **Unit Price** и **Total** – воспользуемся ранее описанными вспомогательными элементами, необходимыми нам для поиска полей **TotalQuantity** и **TotalAmount**.

Во-вторых, Footer таблицы. Он также (как и заголовки столбцов) может быть задан двумя способами: через список ключевых слов и через вспомогательный элемент. Проанализировав ключевые слова итоговой части таблицы, и заметив, что на некоторых изображениях они так же фигурируют в первой строчке таблицы, решаем искать их с помощью отдельного элемента, для которого мы можем задать необходимые ограничения на область поиска.

В-третьих, область поиска таблицы. Справившись с ограничением таблицы сверху (Header) и снизу (Footer), нам нужно ограничить таблицу справа (потребности в ограничении слева нет, т.к. с этой стороны на изображениях нет других данных). Т.к. даже с помощью контекста данных столбца мы не сможем отделить цифры столбца **Total** от цифр столбца **Sales**, нам потребуется другой подход. Можно ограничить область поиска таблицы справа относительно заголовка последнего, не нужного нам, столбца **Sales**.

Замечание. Выбор элементов и последовательности их поиска является творческим процессом и может меняться в процессе отладки гибкого описания.

Таким образом, прежде чем непосредственно приступить к поиску оставшихся полей, опишем все вспомогательные элементы.

Проанализировав взаимозависимости планируемых элементов, начнем с заголовков столбцов таблицы. Нам нужно найти заголовки для столбцов **Quantity**, **Unit Price**, **Total** и **Sales**.

Проанализировав наши изображения, видим, что шапка таблицы начинается с горизонтального сепаратора, и других сепараторов между заголовками столбцов и полями **Invoice Number**, **Invoice Date** и **Delivery Address**.

Данный сепаратор позволит нам очень строго задать область поиска заголовков столбцов таблицы:

1. Для этого нам необходим элемент **hsTableHeaderTop** (тип **Separator**), (подробная инструкция в Шаге 14). Заголовки столбцов мы будем искать в порядке их следования.
2. Создадим составной элемент **TableHeader** (тип **Group**), (подробная инструкция в Шаге 15). В составе которого опишем элементы:
3. элемент **kwQuantity** (тип **Static Text**), соответствующий столбцу **Quantity** таблицы **InvoiceTable** (подробная инструкция в Шаге 16).
4. элемент **kwUnitPrice** (тип **Static Text**), соответствующий столбцу **Unit Price** таблицы **InvoiceTable** (подробная инструкция в Шаге 17).
5. элемент **kwTotal** (тип **Static Text**), соответствующий столбцу **Total** таблицы **InvoiceTable** (подробная инструкция в Шаге 18).
6. элемент **kwSales** (тип **Static Text**), соответствующий столбцу **Sales** таблицы **InvoiceTable** (подробная инструкция в Шаге 19).
В нижней части страницы присутствуют заголовки полей **Total Quantity** и **Total Amount** (по совместительству являющийся для таблицы ограничителем снизу), нас также интересуют альтернативные надписи, которые также могут служить ограничителями для таблицы снизу, и в этой же логической группе опишем заголовки и source element для поля **Country**.
7. Создадим составной элемент **Footer** (тип **Group**).
В составе которого опишем элементы:
8. элемент **kwFooter** (тип **Static Text**), соответствующий итоговой части таблицы **Invoice Table** (подробная инструкция в Шаге 21).
9. элемент **kwTotal** (тип **Static Text**), соответствующий заголовку полей **Total Quantity** и **Total Amount** (подробная инструкция в Шаге 22).
10. элемент **kwOrigin** (тип **Static Text**), соответствующий заголовку поля **Country** (подробная инструкция в Шаге 23).
11. элемент **Country** (тип **Character String**), соответствующий полю **Country** (подробная инструкция в Шаге 24).
Теперь, когда все предварительные приготовления сделаны и описан порядок создания вспомогательных элементов, приступим к анализу поиска оставшихся полей. Опираясь на ранее найденные заголовки столбцов и футера, опишем оставшиеся поля **Total Quantity**, **Total Amount** и **Invoice Table**.
12. Для этого создадим элементы **TotalQuantity** (тип **Character String**) и **TotalAmount** (тип **Character String**), соответствующие полям **Total Quantity** и **Total Amount** (подробная инструкция в Шаге 25).
13. Создадим элемент **InvoiceTable** (тип **Table**), соответствующий полю **Invoice Table** (подробная инструкция в Шаге 26).

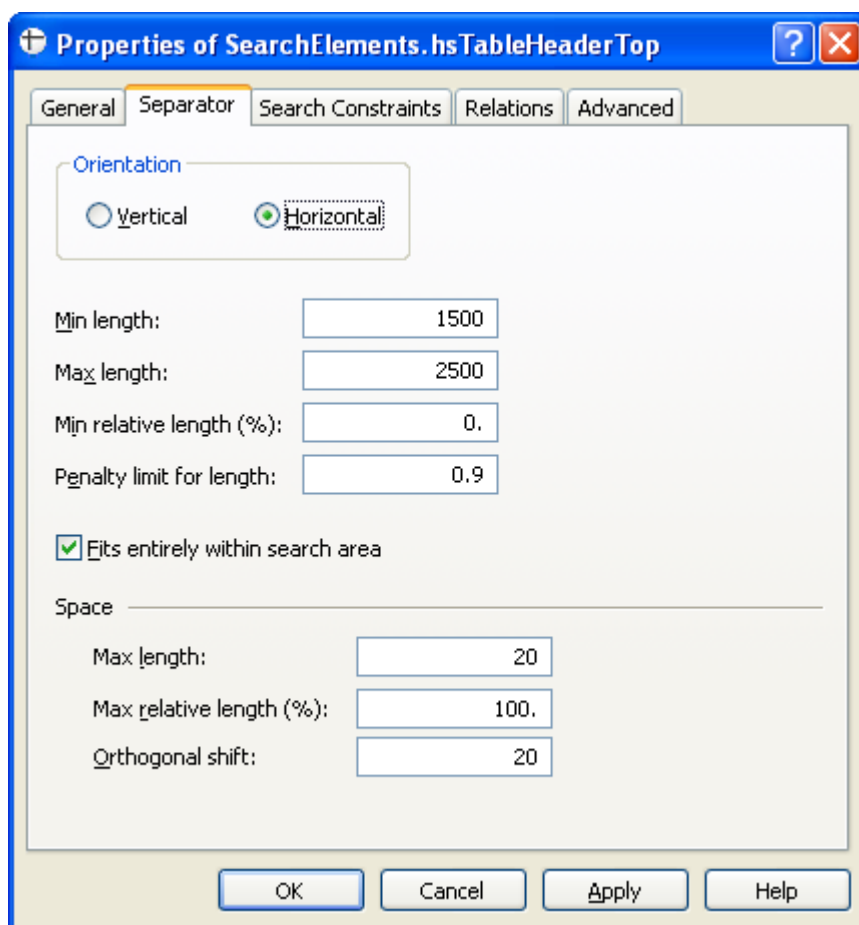
Шаг 14. Поиск вспомогательного горизонтального сепаратора, элемент hsTableHeaderTop

Для того чтобы создать элемент **hsTableHeaderTop**:


1. Создайте элемент **hsTableHeaderTop** (тип **Separator**).
2. Выберите закладку **Separator**.
3. В группе **Orientation** выберите **Horizontal**.
4. Задайте ограничения на длину (**Min/Max length**) от 1500 до 2500.
5. Отметьте опцию **Fit entirely within search area**.
6. Задайте уровень белого мусора (**Space, Max length**) 20, размер ортогонального сдвига (**Orthogonal shift**) 20.

Замечание. При подборе значений для длины сепаратора бывает полезно обратить внимание на геометрические свойства (границы и размеры) близлежащих объектов изображения в режиме просмотра результатов предраспознавания. Значения параметров для порога повреждений сепаратора подбираются опытным путем.

Не накладывайте других ограничений на параметры элемента, оставьте значения по умолчанию.



7. Перейдите на закладку **Relations**.
8. Поскольку сепаратор может находиться только ниже от ключевых слов заголовков **kwInvoiceDate** и **kwInvoiceNumber**, задайте следующее условие поиска:
 - Below элемента InvoiceHeader.kwInvoiceDate, Offset = 0;
 - Below элемента InvoiceHeader.kwInvoiceNumber, Offset = 0;
9. Выберите закладку **Advanced**.
10. Поскольку интересующий нас сепаратор является самым верхним в описанной области поиска, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения ближайший к верху страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:
 - NearestY: PageRect.Top;

 **Замечание.** Данное условие также можно задать через интерфейс программы. Для этого перейдите на закладку **Relations**, отметьте опцию **Nearest** и в выпадающем списке **To**: выберите Page top edge.
11. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях.

Шаг 15. Анализ критериев поиска для заголовков столбцов таблицы. Составной элемент TableHeader

Проанализировав изображения, видим, что заголовки столбцов таблицы располагаются сразу под предварительно найденным горизонтальным сепаратором. Для того чтобы задать настройки, общие для поиска всех заголовков столбцов таблицы, нам потребуется составной элемент.

Для того чтобы создать элемент **TableHeader**:

1. Создайте элемент **TableHeader** (тип **Group**).
2. Выберите закладку **Advanced**.
3. Поскольку при наличии сепаратора на изображении заголовки столбцов таблицы находятся ниже сепаратора **hsTableHeaderTop**, но не ниже чем 100 dot от него, а иначе в верхних трех пятых страницы, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Если элемент **hsTableHeaderTop** найден, то искать объекты изображения ниже сепаратора, но не более чем на 100 dot, иначе искать объекты изображения в верхних трех пятых страницы.* На языке описания FlexiLayout данное условие выглядит следующим образом:

```

If Not (hsTableHeaderTop.IsNull) Then
{
Below: hsTableHeaderTop, 0 * dot;
Above: hsTableHeaderTop.Bottom, -100 * dot;
} Else
{
Above: PageRect.Top + PageRect.Height*3/5;
}

```

Шаг 16. Поиск заголовка столбца Quantity, элемент kwQuantity

Проанализировав результаты предраспознавания, видим что, хотя в данном случае заголовок представлен очень коротким ключевым словом «Qty», качество изображений высокое, и мы можем искать заголовков с помощью элемента типа **Static Text**. Зададим допустимый процент ошибок равным 35%, что для слова из трех букв означает ошибку в одной букве.

Для того чтобы создать элемент **kwQuantity**:

1. В составе элемента **TableHeader** создайте элемент **kwQuantity** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **Qty**
4. Задайте значение свойства **Max error percentage** равным 35.
5. Выберите закладку **Advanced**.
6. Поскольку заголовок столбца **Quantity** находится правее первой трети изображения по горизонтали, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения правее первой трети страницы.* На языке описания FlexiLayout данное условие выглядит следующим образом:


```
RightOf: PageRect.Left + PageRect.Width/3;
```
7. Протестируйте изображения и убедитесь, что элемент находится на всех изображениях.

Шаг 17. Поиск заголовка столбца UnitPrice, элемент kwUnitPrice

При описании элемента для поиска ключевых слов заголовка столбца **Unit Price** воспользуемся наблюдением, что столбец **Unit Price** всегда располагается правее столбца **Quantity**.

Для того чтобы создать элемент **kwUnitPrice**:

1. В составе элемента **TableHeader** создайте элемент **kwUnitPrice** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **UnitPrice**
4. Выберите закладку **Advanced**.
5. Поскольку заголовок столбца **Unit Price** располагается правее заголовка столбца **kwQuantity**, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Если элемент **kwQuantity** найден, то искать объект изображения правее элемента **kwQuantity**.* На языке описания FlexiLayout данное условие выглядит следующим образом:


```

If Not (TableHeader.kwQuantity.IsNull) Then
RightOf: TableHeader.kwQuantity, 0 * dot;

```
6. Протестируйте изображения и убедитесь, что элемент находится на всех изображениях.

Шаг 18. Поиск заголовка столбца Total, элемент kwTotal

Столбец **Total** на всех изображениях находится правее столбца **Unit Price**. Поэтому будем строить поиск заголовка столбца **Total** относительно заголовка столбца **Unit Price**.

Для того чтобы создать элемент **kwTotal**:

1. В составе элемента **TableHeader** создайте элемент **kwTotal** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **TotalChf**
4. Выберите закладку **Advanced**.
5. Поскольку заголовок столбца **Total** располагается правее заголовка столбца **kwUnitPrice** (если последний находится на изображении), в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Если элемент **kwUnitPrice** найден, то искать объект изображения правее элемента **kwUnitPrice***. На языке описания FlexiLayout данное условие выглядит следующим образом:

```
If Not (TableHeader.kwUnitPrice.IsNull) Then
RightOf: TableHeader.kwUnitPrice, 0 * dot;
```

6. Протестируйте изображения и убедитесь, что элемент находится на всех изображениях.

Шаг 19. Поиск заголовка столбца Sales, элемент kwSales

Хотя сам столбец **Sales** нам не нужен, его заголовок необходим нам для ограничения таблицы справа.

Для того чтобы создать элемент **kwSales**:

1. В составе элемента **TableHeader** создайте элемент **kwSales** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **Sales**
4. Выберите закладку **Advanced**.
5. Поскольку заголовок столбца **Sales** располагается правее заголовка столбца **kwTotal** (если последний находится на изображении), в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Если элемент **kwTotal** найден, то искать объект изображения правее элемента **kwTotal***. На языке описания FlexiLayout данное условие выглядит следующим образом:

```
If Not (TableHeader.kwTotal.IsNull) Then
RightOf: TableHeader.kwTotal, 0 * dot;
```

6. Протестируйте изображения и убедитесь, что элемент находится на всех изображениях.

Шаг 20. Описание группы Footer, элемент Footer

Проанализировав изображения, видим, что футер таблицы, поля **Country**, **Total Quantity** и **Total Amount**, а так же их заголовки, располагаются в нижней части изображения, а именно под заголовочной областью таблицы. Чтобы не захватить первую строку таблицы, в которой могут дублироваться ключевые слова итоговой части, сместим верхнюю границу области поиска на небольшой отступ вниз. Для того чтобы задать настройки, общие для поиска всех этих элементов, нам потребуется составной элемент.

Для того чтобы создать элемент **Footer**:

1. Создайте элемент **Footer** (тип **Group**).
2. Перейдите на закладку **Relations**.
3. Поскольку итоговая часть таблицы **Invoice Table** и поля **Country**, **Total Quantity** и **Total Amount** могут находиться только ниже сепаратора **hsTableHeaderTop** с некоторым запасом, задайте следующее условие поиска:

– **Below** элемента **hsTableHeaderTop**, **Offset** = 300

Шаг 21. Описание итоговой части таблицы, элемент kwFooter

Изучив изображения, видим, что итоговая часть таблицы может начинаться с ключевых слов «Carried Over», «Total Chf», «Goods total excl. tax», «Discount» и «SPECIAL DISCOUNT». Включим все эти варианты в поисковый текст элемента типа **Static Text**.

Для того чтобы создать элемент **kwFooter**:

1. В составе элемента **Footer** создайте элемент **kwFooter** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста:
Discount|TotalCHF|SPECIALDISCOUNT|Carriedover|Goodstotalexcl.tax
4. Выберите закладку **Advanced**.
5. Поскольку искомые ключевые слова находятся в левой половине изображения, и если их будет несколько, интересующий нас случай будет самым верхним, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения левее середины страницы. Искать объект изображения ближайший к верху страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:

LeftOf: PageRect.Left + PageRect.Width/2;
NearestY: PageRect.Top;
6. Протестируйте изображения и убедитесь, что элемент находится на всех изображениях, где он присутствует.

Замечание. В реальной практике подобная ситуация, когда текст надписи может быть представлен более чем одним вариантом, является сигналом: нужно убедиться, что учтены все варианты написания заголовка, или хотя бы подавляющее большинство их. Можно запросить больше тестовых изображений или постараться узнать у клиента каким-то другим способом, какие варианты написания заголовка поля могут присутствовать на данном типе форм.

Шаг 22. Описание заголовка полей Total, элемент kwTotal

Проанализировав изображения, видим, что заголовок полей **Total Quantity** и **Total Amount** представлен ключевыми словами «TotalChf» или «Carried Over» и находится в левой нижней части страницы.

Для того чтобы создать элемент **kwTotal**:

1. В составе элемента **Footer** создайте элемент **kwTotal** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **TotalChf|CarriedOver**
4. Задайте значение свойства **Max error percentage** равным 20.
5. Выберите закладку **Advanced**.
6. Поскольку данный заголовок находится в левой половине изображения по горизонтали, а так же интересующая нас строка располагается самой верхней, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения левее середины страницы. Искать объект изображения ближний к верху страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:

LeftOf: PageRect.Left + PageRect.Width/2;
NearestY: PageRect.Top;
7. Протестируйте изображения и убедитесь, что элемент находится на всех изображениях, где он присутствует.

Шаг 23. Поиск заголовка поля Country, элемент kwOrigin

Для поиска заголовка поля **Country** по ключевым словам «OriginOfGoods» воспользуемся элементом типа **Static Text**.

Для того чтобы создать элемент **kwOrigin**:

1. В составе элемента **Footer** создайте элемент **kwOrigin** (тип **Static Text**).
2. Выберите закладку **Static Text**.
3. В поле **Search Text** укажите значение искомого текста: **OriginOfGoods**:
4. Перейдите на закладку **Relations**.

5. Поскольку поле **Country** может находиться только ниже итоговой части таблицы **kwFooter**, задайте следующее условие поиска:
 - **Below** элемента **Footer.kwFooter**, **Offset = 0**
6. Выберите закладку **Advanced**.
7. Поскольку поле **Country** находится в левой половине изображения, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения в левой половине страницы*. На языке описания FlexiLayout данное условие выглядит следующим образом:
 - LeftOf: PageRect.Left + PageRect.Width/2;
8. Протестируйте изображения и убедитесь, что элемент находится на всех изображениях, где он присутствует.

Шаг 24. Описание поля Country, элемент Country

Поле **Country** мы будем искать привычным способом с помощью элемента типа **Character String** на одном уровне с заголовком этого поля и только в том случае, если заголовок на данном изображении найден.

Для того чтобы создать элемент **Country**:


1. В составе элемента **Footer** создайте элемент **Country** (тип **Character String**).
2. Выберите закладку **Character String**.
3. Задайте алфавит допустимых символов из всех латинских букв, для неалфавитных символов поставьте ограничение в 20 %.
4. В поле **String Length** задайте нечеткий интервал { -1, 2, 25, INF }, оценивающий длину цепочки символов. Предполагаем, что все возможные значения будут находиться в интервале от 2 до 25 символов. Любая гипотеза, выходящая за рамки данного интервала, будет штрафоваться.
5. Не накладывайте других ограничений на параметры элемента, оставьте значения по умолчанию.
6. Перейдите на закладку **Relations**.
7. Поскольку **Country** может находиться только справа от ключевых слов заголовка **kwOrigin** с небольшим отступом, но не дальше, чем 500 dot от правого края заголовка, задайте следующее условие поиска:
 - **RightOf** элемента **kwOrigin**, **Offset = 10**
 - **LeftOf** элемента **kwOrigin**, **Offset = -500**, **Element border = Right**
8. Поскольку **Country** всегда находится на одном уровне с ключевыми словами заголовка **kwOrigin**, добавьте следующие условия поиска:
 - **Above** элемента **kwOrigin**, **Offset = -10**, **Element border = Bottom**
 - **Below** элемента **kwOrigin**, **Offset = -10**, **Element border = Top**

 **Замечание.** При подборе значений для отступов (offset) бывает полезно обратить внимание на геометрические свойства (границы и размеры) опорных и искомых объектов изображения в режиме просмотра результатов предраспознавания. Подбор значений для отступов (offset) является творческим процессом и осуществляется путем проб и ошибок.
9. Выберите закладку **Advanced**.
10. Поскольку поле **Country** может находиться на изображении, а может и нет, но при этом всегда присутствует вместе с заголовком, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения, только если найден элемент kwOrigin*. На языке описания FlexiLayout данное условие выглядит следующим образом:
 - If Footer.kwOrigin.IsNull Then DontFind;
11. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях, где он присутствует.
12. Выберите в качестве описания местоположения блока **Country** опцию **Source element**. В качестве source element укажите элемент **Country**, нажав кнопку "...".

Шаг 25. Поиск полей TotalQuantity и TotalAmount, элементы TotalQuantity и TotalAmount

Проанализировав тестовые изображения, видим, что интересующие нас поля **Total Quantity** и **Total Amount** присутствуют только вместе с заголовком **Footer.kwTotal** и располагаются на одном с ним уровне. Поля **Total Quantity** и **Total Amount** будем искать с помощью элементов типа **Character String**. Для того чтобы задать настройки, общие для этих элементов, нам потребуется составной элемент.

Для того чтобы создать составной элемент **grTotal**:

1. В составе элемента **Footer** создайте элемент **grTotal** (тип **Group**).
2. Отметьте не заданное по умолчанию свойство **Optional** на закладке **General** (поскольку мы решили не искать элементы данной группы в случае отсутствия заголовка поля на форме).
 **Замечание.** Особенности работы с опциональным элементом типа Group см. в разделе справки *Советы и рекомендации* \Свойство Optional составного элемента.
3. Выберите закладку **Advanced**.
4. Поскольку поля **Total Quantity** и **Total Amount** присутствуют на изображении только вместе с заголовком **Footer.kwTotal**, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Если элемент Footer.kwTotal не найден, то не искать объекты изображения*. На языке описания FlexiLayout данное условие выглядит следующим образом:

```
If Footer.kwTotal.IsNull Then DontFind;
```


Поле **Total Quantity** располагается на пересечении строки с заголовком **Total** и столбца **Quantity**. Первое условие уже задано на уровне группы, а для задания второго условия мы воспользуемся предварительно найденным заголовком столбца в качестве опорного элемента (reference element).

Для того чтобы создать элемент **TotalQuantity**:

1. В составе элемента **grTotal** создайте элемент **TotalQuantity** (тип **Character String**).
2. Выберите закладку **Character String**.
3. В поле **String Length** задайте нечеткий интервал { -1, 1, 10, INF }, оценивающий длину цепочки символов. Предполагаем, что все возможные значения будут находиться в интервале от 1 до 10 символов. Любая гипотеза, выходящая за рамки данного интервала, будет штрафоваться.
4. Не накладывайте других ограничений на параметры элемента, оставьте значения по умолчанию.
5. Перейдите на закладку **Relations**.
6. Поскольку поле **Total Quantity** всегда находится на одном уровне с ключевыми словами заголовка **Footer.kwTotal**, либо несколько ниже, добавьте следующие условия поиска:

– **Above** элемента **Footer.kwTotal**, **Offset** = -200, **Element border** = Bottom

– **Below** элемента **Footer.kwTotal**, **Offset** = -20, **Element border** = Top

 **Замечание.** При подборе значений для отступов (offset) бывает полезно обратить внимание на геометрические свойства (границы и размеры) опорных и искомых объектов изображения в режиме просмотра результатов предраспознавания. Подбор значений для отступов (offset) является творческим процессом и осуществляется путем проб и ошибок.

7. Выберите закладку **Advanced**.
8. Поскольку поле **Total Quantity** может находиться на изображении, а может и нет, но при этом всегда присутствует только вместе со столбцом **Quantity**, под заголовком столбца с небольшим запасом, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Искать объект изображения, только если найден элемент TableHeader.kwQuantity, при этом искать не левее чем 50 dot от левой границы заголовка и не правее чем 50 dot от правой границы заголовка столбца*. На языке описания FlexiLayout данное условие выглядит следующим образом:

```
If Not (TableHeader.kwQuantity.IsNull) Then
{
LeftOf: TableHeader.kwQuantity.Right, -50 * dot;
RightOf: TableHeader.kwQuantity.Left, -50 * dot;
}
Else DontFind;
```


9. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях, где он присутствует.
10. Выберите в качестве описания местоположения блока **TotalQuantity** опцию **Source element**. В качестве source element укажите элемент **TotalQuantity**, нажав кнопку "...".

Поле **Total Amount** располагается на пересечении строки с заголовком **Total** (футера таблицы) и столбца таблицы **Total**. Напомним, что первое условие уже задано на уровне группы. Для задания второго условия мы воспользуемся предварительно найденными заголовками столбцов в качестве опорного элемента (reference element).

Для того чтобы создать элемент **TotalAmount**:

1. В составе элемента **grTotal** создайте элемент **TotalAmount** (тип **Character String**).
2. Выберите закладку **Character String**.
3. В поле **String Length** задайте нечеткий интервал { -1, 1, 20, INF }, оценивающий длину цепочки символов. Предполагаем, что все возможные значения будут находиться в интервале от 1 до 20 символов. Любая гипотеза, выходящая за рамки данного интервала, будет штрафовать.
4. Не накладывайте других ограничений на параметры элемента, оставьте значения по умолчанию.
5. Перейдите на закладку **Relations**.
6. Поскольку поле **Total Amount** всегда находится на одном уровне с ключевыми словами заголовка **Footer.kwTotal**, добавьте следующие условия поиска:
 - **Above** элемента **Footer.kwTotal**, **Offset** = -20, **Element border** = Bottom
 - **Below** элемента **Footer.kwTotal**, **Offset** = -20, **Element border** = Top
7. Выберите закладку **Advanced**.
8. Поскольку поле **Total Amount** может находиться только непосредственно под заголовком столбца **Total** (с некоторым запасом) по вертикали, другими словами между заголовков столбцов **UnitPrice** и **Sales** по вертикали, тоже с небольшим запасом, в поле **Advanced pre-search relations** задайте дополнительное условие поиска: *Если найден элемент **TableHeader.kwTotal**, то искать объект изображения не правее чем 70 dot от правой границы элемента **TableHeader.kwTotal** и не левее чем 50 dot левой границы элемента **TableHeader.kwTotal**; иначе если найдены элементы **TableHeader.kwUnitPrice** и **TableHeader.kwSales**, то искать объект изображения не правее чем 40 dot от левой границы элемента **TableHeader.kwSales** и правее правой границы элемента **TableHeader.kwUnitPrice**; иначе объект изображения не искать.* На языке описания FlexiLayout данное условие выглядит следующим образом:

```
If Not (TableHeader.kwTotal.IsNull) Then
{
LeftOf: TableHeader.kwTotal.Right, -70 * dot;
RightOf: TableHeader.kwTotal.Left, -50 * dot;
}
Else
If Not (TableHeader.kwUnitPrice.IsNull) and Not (TableHeader.kwSales.IsNull) Then
{
LeftOf: TableHeader.kwSales.Left, -40 * dot;
RightOf: TableHeader.kwUnitPrice, 0 * dot;
}
Else DontFind;
```

 **Замечание.** При подборе значений для отступов (offset) бывает полезно обратить внимание на геометрические свойства (границы и размеры) опорных и искомым объектов изображения в режиме просмотра результатов предраспознавания. Подбор значений для отступов (offset) является творческим процессом и осуществляется путем проб и ошибок.

9. Протестируйте изображения и убедитесь, что программа верно находит элемент на всех изображениях, где он присутствует.
10. Выберите в качестве описания местоположения блока **TotalAmount** опцию **Source element**. В качестве source element укажите элемент **TotalAmount**, нажав кнопку "...".

Шаг 26. Поиск табличного элемента, элемент InvoiceTable

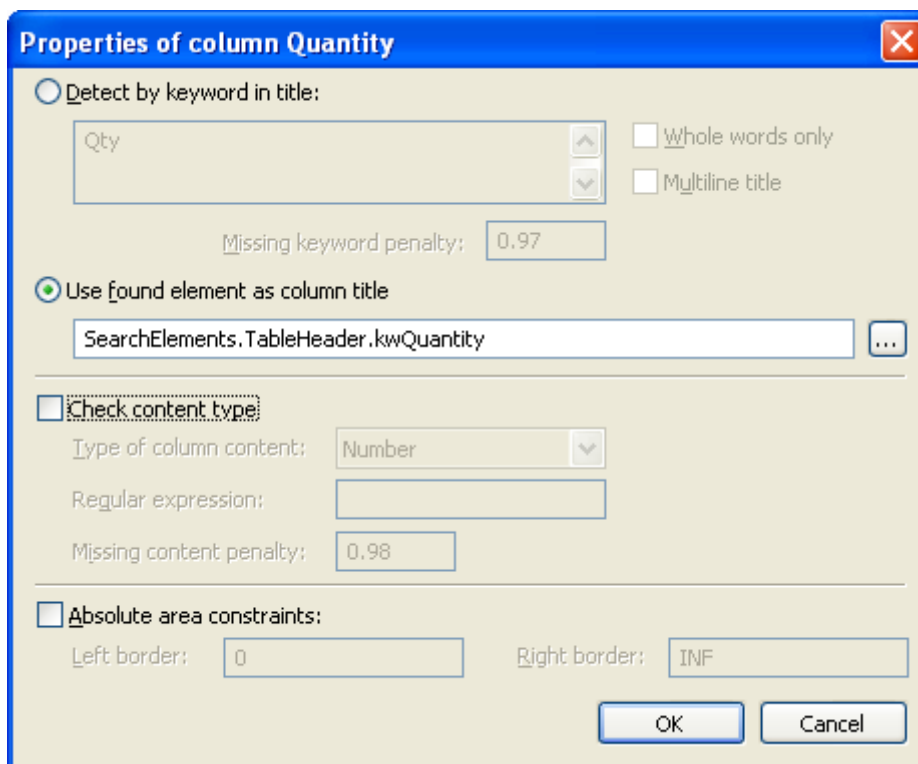
Итак, все предварительные приготовления произведены, и мы можем искать поле **Invoice Table**. Опишем элемент типа **Table**. Во-первых, Header таблицы. Как уже было решено (подробнее Шаг 13), заголовки первых 2х столбцов **Reference** и **Designation** мы опишем с помощью соответствующих ключевых слов (диалог свойств табличного элемента -> закладка **Columns** -> диалог свойств столбца), а для остальных - столбцов **Quantity**, **Unit Price** и **Total** - воспользуемся предварительно описанными вспомогательными элементами.

Во-вторых, Footer таблицы. Для задания футера таблицы (Footer) мы решили (см. подробнее Шаг 13) использовать предварительно описанный элемент, отыскивающий начало итоговой части по ключевым словам и фразам.

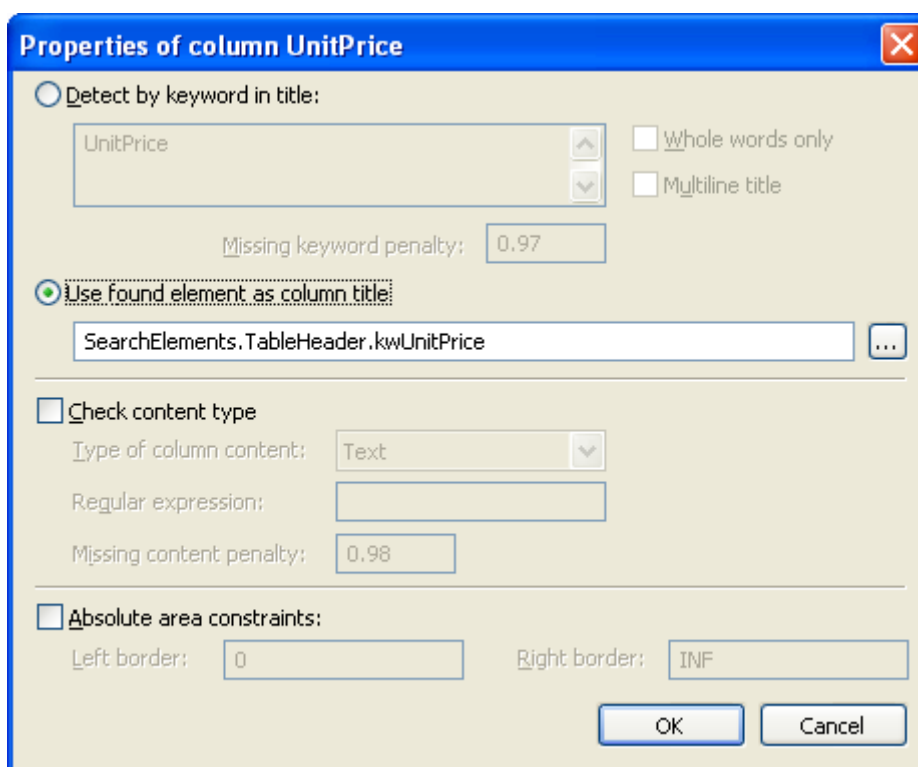
В-третьих, область поиска таблицы. Ограничим область поиска справа относительно заголовка последнего, не нужного нам, столбца **Sales**.

Для того чтобы создать элемент **InvoiceTable**:

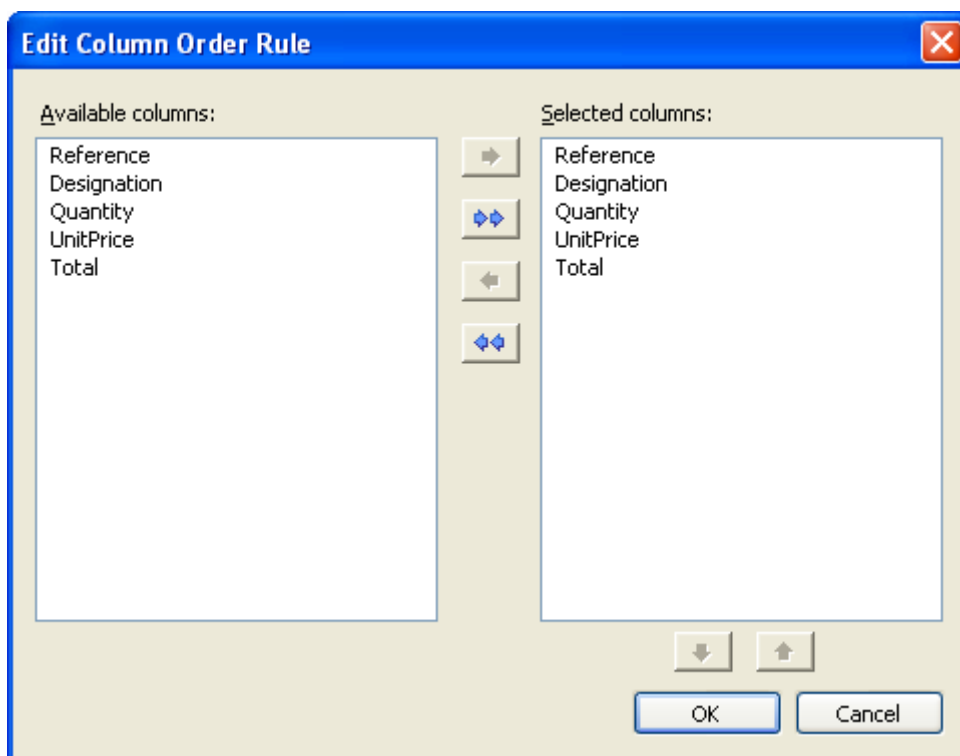
1. Создайте элемент **InvoiceTable** (тип **Table**).
2. Выберите закладку **Columns**.
3. Выберите в качестве блока, для поиска которого служит данный элемент, ранее созданный (см. Шаг 3) блок **InvoiceTable**, нажав кнопку "...".
4. Выберите в списке столбцов **Reference** и нажмите кнопку **Properties**. Заголовок данного столбца мы будем искать по ключевому слову **Reference**, поэтому оставьте значения по умолчанию.
5. Выберите в списке столбцов **Designation** и нажмите кнопку **Properties**. Заголовок данного столбца мы будем искать по ключевому слову **Designation**, поэтому оставьте значения по умолчанию.
6. Выберите в списке столбцов **Quantity** и нажмите кнопку **Properties**. Для заголовка данного столбца мы будем использовать ранее найденный элемент **TableHeader.kwQuantity**. Для этого переключитесь на опцию **Use found element as column title** и выберите элемент **TableHeader.kwQuantity**, нажав кнопку "...".

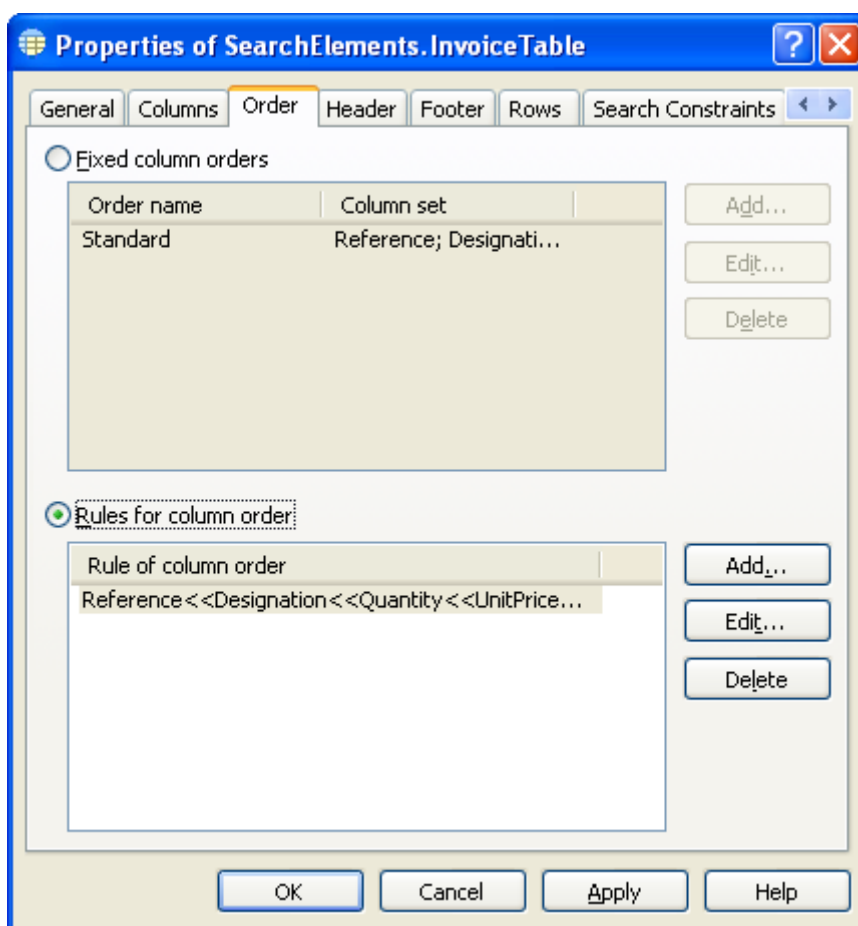


7. Выберите в списке столбцов **UnitPrice** и нажмите кнопку **Properties**. Для заголовка данного столбца мы будем использовать ранее найденный элемент **TableHeader.kwUnitPrice**. Для этого переключитесь на опцию **Use found element as column title** и выберите элемент **TableHeader.kwUnitPrice**, нажав кнопку "...".

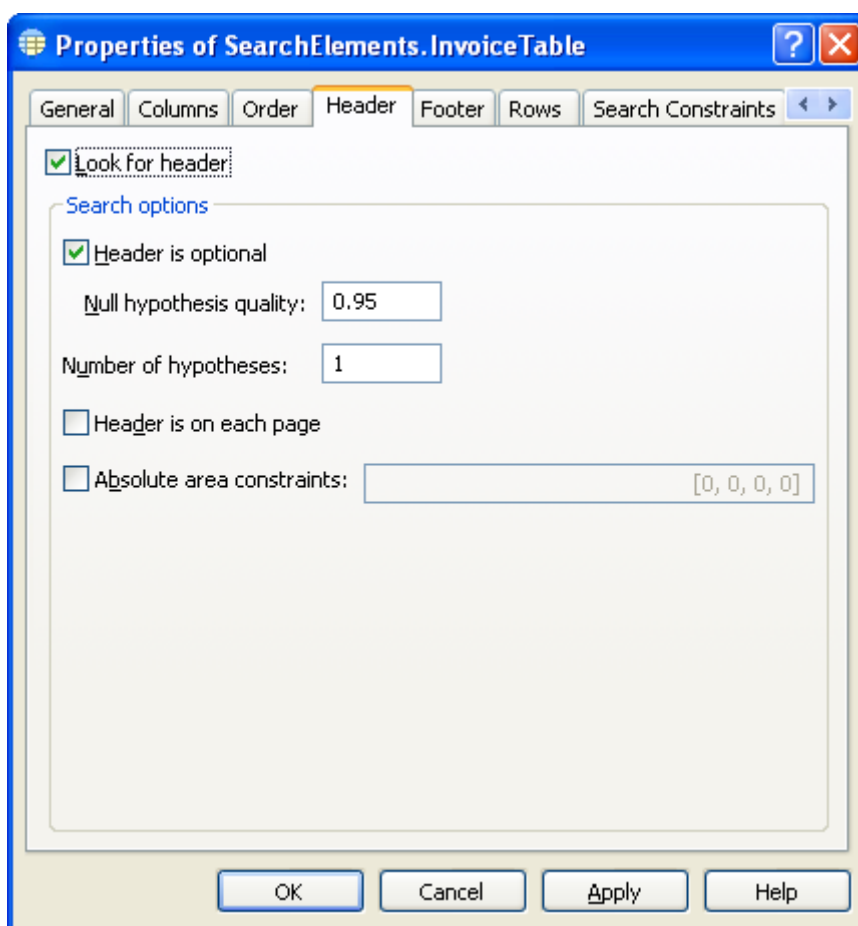


8. Выберите в списке столбцов **Total** и нажмите кнопку **Properties**. Для заголовка данного столбца мы будем использовать ранее найденный элемент **TableHeader.kwTotal**. Для этого переключитесь на опцию **Use found element as column title** и выберите элемент **TableHeader.kwTotal**, нажав кнопку "...".
9. Выберите закладку **Order**.
10. Поскольку искомые столбцы могут перемежаться другими (например, **Unit**), но всегда следуют в следующем порядке: сначала **Reference**, потом **Designation**, потом **Quantity**, правее него **Unit Price**, и ещё правее **Total**, – то выберем вариант **Rules for column order** и, нажав кнопку "Add...", добавим все столбцы в таком порядке: **Reference << Designation << Quantity << Unit Price << Total**.

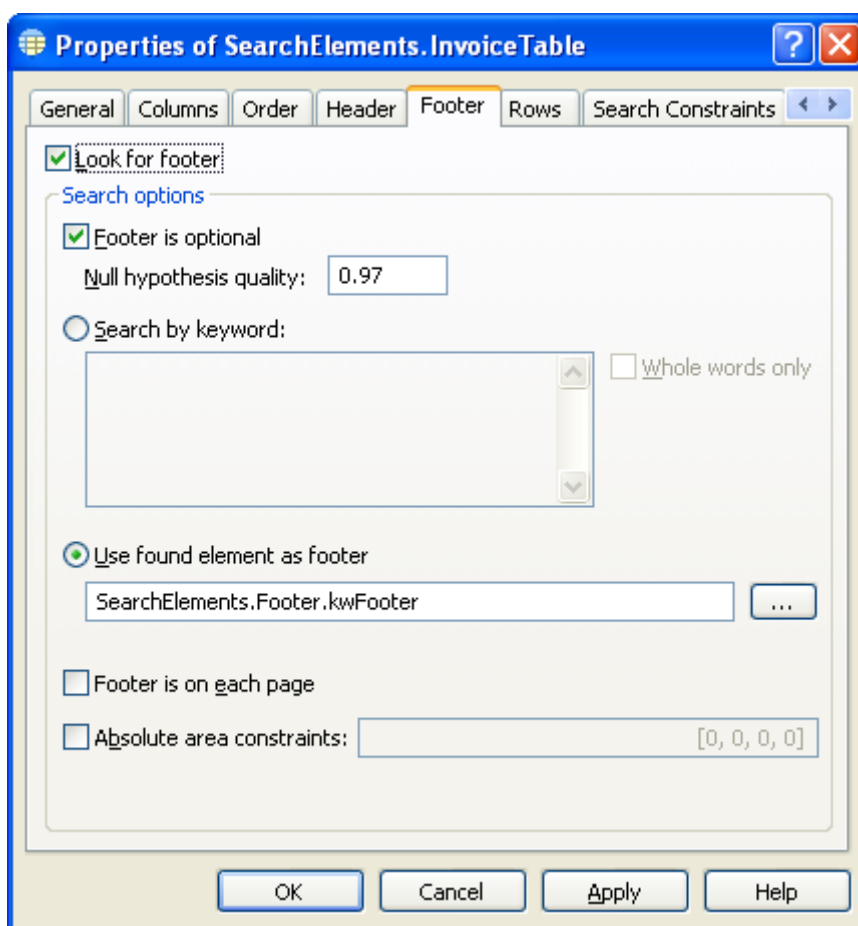




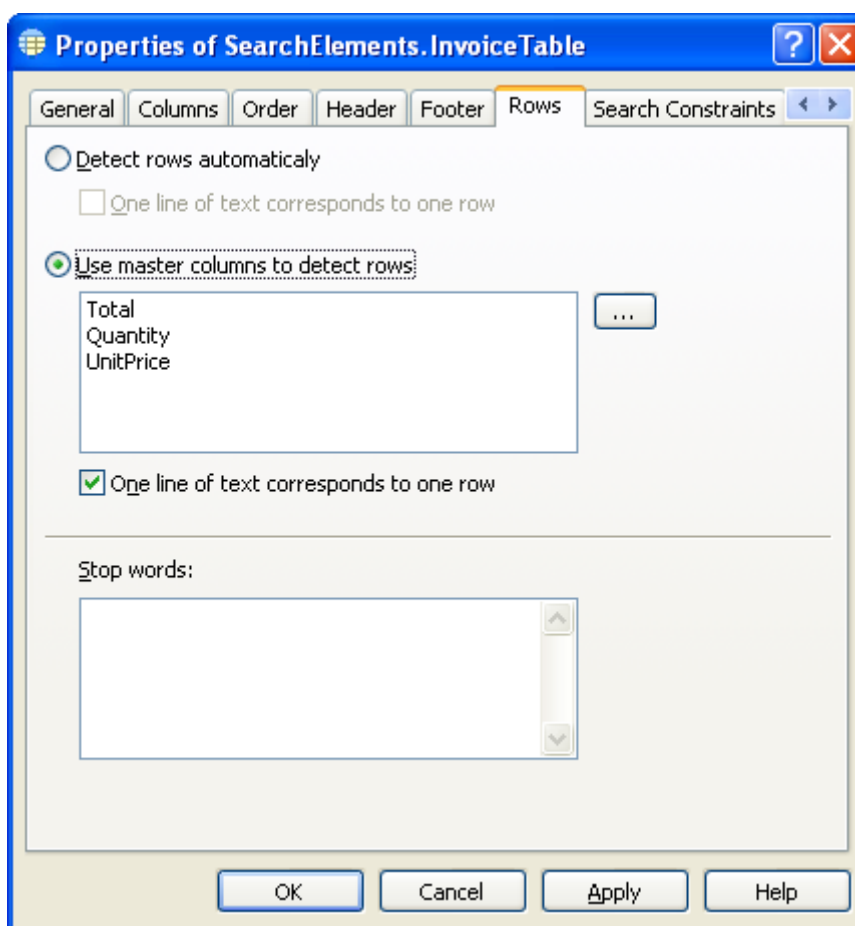
11. Выберите закладку **Header**. Заголовок таблицы нам нужен, но других ограничений мы накладывать не будем, поэтому оставьте значения по умолчанию.



12. Выберите закладку **Footer**. Для итоговой части таблицы мы будем использовать ранее найденный элемент **Footer.kwFooter**. Для этого переключитесь на опцию **Use found element as footer** и выберите элемент **Footer.kwFooter**, нажав кнопку "...".

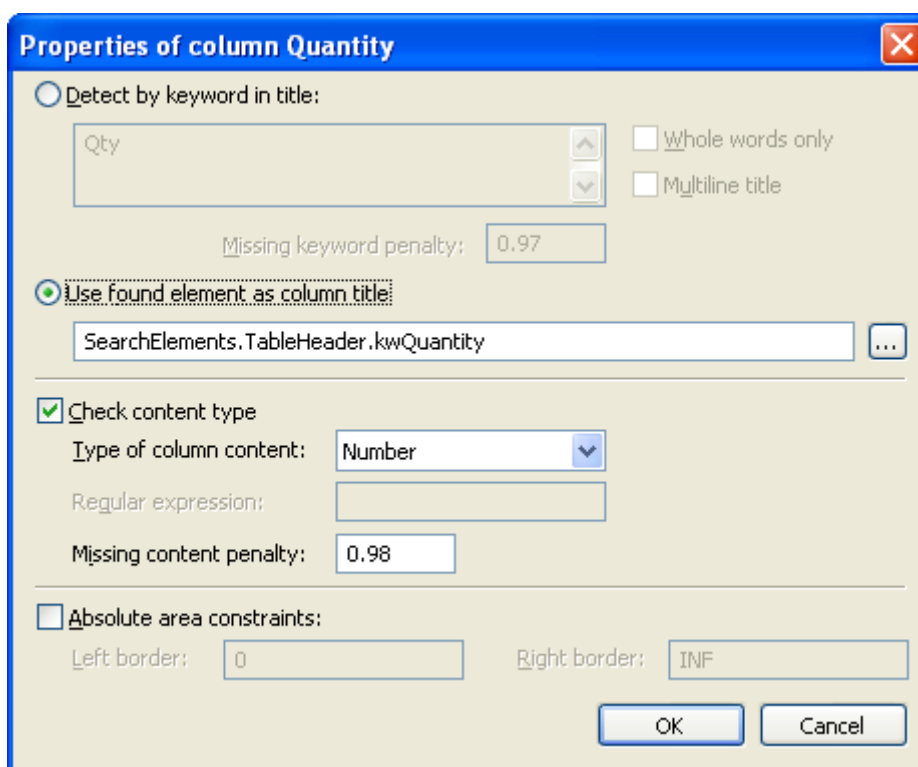


13. Выберите закладку **Rows**.
14. Поскольку в нашей таблице некоторые столбцы могут содержать многострочные (multiline) записи (rows), необходимо задать ключевые столбцы для разделения на строки. Выберите опцию **Use master columns to detect rows**. В качестве разделителя на строки первым укажем столбец **Total**, т.к. в этом столбце всегда присутствуют данные, и каждой записи таблицы в этом столбце соответствует одна строка данных. На тот случай, если данный столбец не будет обнаружен на форме, укажем следующими разделителями столбец **Quantity** и столбец **Unit Price**. Выберите данный набор столбцов с помощью кнопки "...".



15. Выберите закладку **Relations**. Поскольку интересующие нас столбцы таблицы располагаются левее начала заголовка столбца **Sales** с некоторым запасом, добавьте следующие условия поиска:
 - **LeftOf** элемента **TableHeader.kwSales**, **Offset** = -40, **Element border** = Left

Замечание. При подборе значений для отступов (offset) бывает полезно обратить внимание на геометрические свойства (границы и размеры) опорных и искомых объектов изображения в режиме просмотра результатов предраспознавания. В остальном эти значения подбираются опытным путем.
16. Наложим шаблон на тестовый набор изображений. Видим, что программа верно находит табличный элемент на всех изображениях, но на некоторых изображениях в столбец **Quantity** попадает текст из предыдущего, не нужного нам столбца. Чтобы отделить данные столбца **Quantity** от мусора, зададим тип контекста для данного столбца.
17. Откройте диалог **Properties** табличного элемента **InvoiceTable**. Перейдите на закладку **Columns**, откройте диалог **Properties** для столбца **Quantity**, отметьте опцию проверки содержимого столбца **Check content type**. В типе контекста **Type of column content** выберите **Number**.



18. Если мы протестируем изображения теперь, то видим, что сейчас в столбец **Quantity** попадают только цифры, что и требовалось.
19. Выберите в качестве описания местоположения блока **InvoiceTable** опцию **Source element**. В качестве source element укажите элемент **InvoiceTable**, нажав кнопку "...".

Шаг 27. Экспорт гибкого описания и подключение его к шаблону FlexiCapture

Полученное гибкое описание можно экспортировать в формат ABBYY FlexiLayout (*.afl) и подключить к шаблону в программе ABBYY FlexiCapture.

 **Замечание.** Подробнее об экспорте и подключении смотрите Пример 1, Шаг 21 – Шаг 22.