FlexiLayout Studio 9.0

Tutorial

Sample 4

Contents

Multi–Page Invoice

Important! A multi–page document is used in this sample.

The sample document described in this section is a typical multi–page invoice. Unlike the sample documents described in the earlier sections, it has more than one page and contains a table that stretches over several pages.

Note: ABBYY FlexiLayout Studio 9.0 allows you to create FlexiLayouts for multi–page documents. Multi–page documents are treated as whole units and search for elements is not limited to individual pages. For details, see Help> Multi–Page FlexiLayout.

The **Invoice** sample document will be used to illustrate some non–standard approaches to locating elements. You will learn how to detect image objects that correspond to the beginning and the end of a multi–page document and how to detect multi–page elements (in this case, a multi–page table). For details, see Help>Elements\Elements Properties\Header and Footer.
A FlexiLayout for the Invoice sample document should reliably detect the following fields on all the test images:

Invoice Number

Invoice Date

Delivery Address

Total Amount

Invoice Table (**Reference**, **Designation**, **Unit**, **Quantity**, **Unit Price**, and **Total** columns)

Company

Step 1. Preliminary settings

1. Run ABBYY FlexiLayout Studio 9.0.
2. Create a new project (**File**>**New Project...**).
3. In the dialog box that opens, type this name for your project: **Invoices**.
4. Add the images to the batch (see. Sample 1. Step 2 for detailed instructions).
   Note: The test images for Sample 4 can be found in **<disk name>:\\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\9.0\Samples\FlexiLayout Studio\Invoice\InvoicesBatch**.
5. For each document in the batch, you can create a reference assembly which shows how the pages should be separated into documents. A reference assembly can be created either manually or automatically, when adding the images to the batch. A reference assembly allows you to compare the actually assembled documents with the correct "ideals." For details, see Help>Testing and Adjusting the FlexiLayout\Reference document assembly.

To correctly assemble pages into documents, you first have to analyze them to see which pages go into to which document and in what order. In the **Batch** window, the pages that belong to the same batch must strictly follow one another (page 1, page 2, page 3, etc.).

Note: If the pages have been added to a batch in a random order and cannot be assembled into documents, they should be renumbered. To renumber the pages, use your mouse to drag and drop each page into the desired location in the **Batch** window. The program will automatically renumber the rearranged pages.

In the **Batch** window, select the pages relating to the same document.

Right–click the selection and select **Assemble to Reference Documents...**.

In the **Assemble to Reference Documents** dialog box, select **One document**. The program will group all the selected pages into one document.

Select **First page of document matches with header**. This option means that the Header element must be matched with the first page of the document.

Select **Last page of document matches with footer**. This option means that the **Footer** element must be matched with the last page of the document.

Click **OK**. The reference layout will be displayed in **Reference Layout** mode: the beginning and end of the document are marked with green markers.

Note: If the pages have been assembled into a document incorrectly, you can cancel the operation by selecting **Disassemble Reference Documents**.

6. Specify the properties of your FlexiLayout:

Name: Invoice.

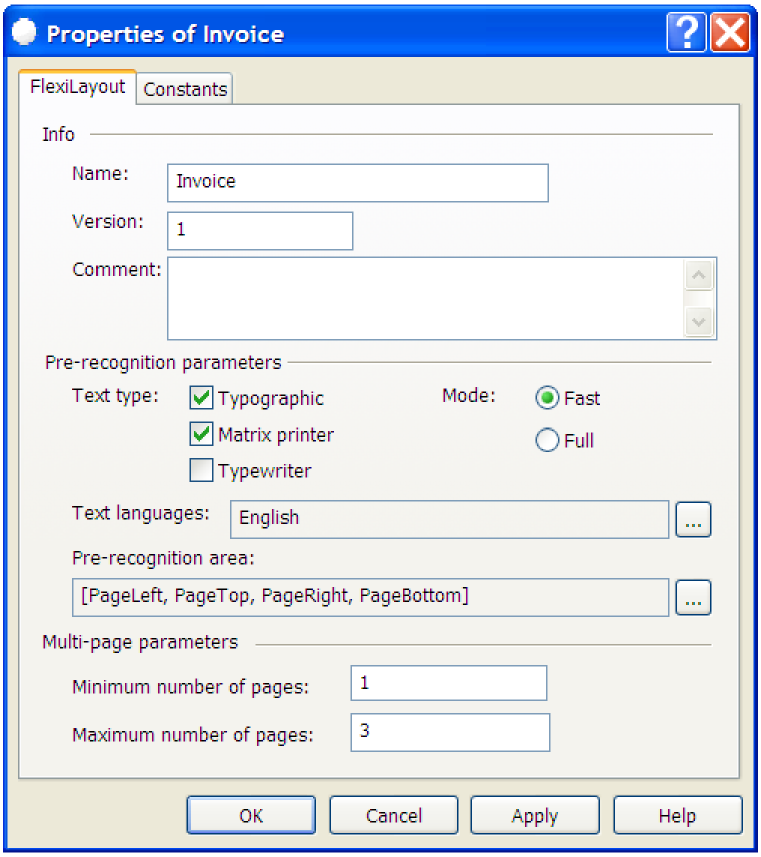
Text Type: select **Typographic** and **Matrix printer** text types.

Text languages: select **English**, as the sample documents are in English.

Mode: select **Fast** recognition mode.

Minimal number of pages: 1, as the sample documents may include a minimum of one pages.

Maximal number of pages: 3, as the sample documents may include a maximum of 3 pages.



Info

Name: Invoice

Version: 1

Comment:

Pre-recognition parameters

Text type: ☑ Typographic ☑ Matrix printer □ Typewriter

Mode: ● Fast ○ Full

Text languages: English

Pre-recognition area: [PageLeft, PageTop, PageRight, PageBottom]

Multi-page parameters

Minimum number of pages: 1

Maximum number of pages: 3

Note: For detailed instructions on selecting pre–recognition parameters, see Sample 1. Step 3.

Step 2. Viewing the images and pre–recognition results

Before you start creating your FlexiLayout, you must consider which of the image objects can be used to identify the beginning and the end of the document and to find the required data fields. You can rely only on those graphical and/or text objects which are consistently detected on the images during pre–recognition.

Note: Special **Header** and **Footer** elements can be used to detect the beginning and the end of a document. The may consist of one or more elements that unambiguously identify the document's beginning and end respectively.

To launch the pre–recognition process:

1. Select all the images in the batch.
2. Select the **Prerecognize** command (**Batch** menu or the shortcut menu of the image).

Note: For more information about the pre–recognition process, see. Sample 1. Step 4.
Once you examine the pre–recognition results, you will see that all the text objects corresponding to the field titles, column titles and horizontal separators are detected reliably and can be used as starting points to look for other objects on the images.

Note: For more information about viewing images and pre–recognition results, see Sample 1. Step 5 and Sample 1. Step 6.
Now you can create and set up reference elements and search constraints for blocks corresponding to data fields. Activate the **FlexiLayout** window and go to the next step.

Step 3. Blocks

Our FlexiLayout should reliably detect the following blocks (provided they are present on a given image):

Invoice Number

Invoice Date

Delivery Address

Total Amount

Invoice Table (**Reference**, **Designation**, **Unit**, **Quantity**, **Unit Price**, and **Total** columns)

Company

For convenience, create a list of all the required blocks:

1. For the **Invoice Number** field, create a block of the **Text** type and name it **InvoiceNumber**.
2. For the **Invoice Date** field, create a block of the **Text** type and name it **InvoiceDate**.
3. For the **Delivery Address** field, create a block of the **Text** type and name it **DeliveryAddress**.
4. For the **Total Amount** field, create a block of the **Text** type and name it **TotalAmount**.
5. For the **Company** field, create a block of the **Text** type and name it **Company**.
6. For the **Invoice Table** field, create a block of the **Table** type and name it **InvoiceTable**.
   In the block's properties dialog box, click the **Columns** tab and add the titles of the columns in the following order: **Reference**, **Designation**, **Unit**, **Quantity**, **Unit Price**, **Total**.

   Note: The provided column titles and their order will correspond to this block in the FlexiCapture template.

Note: For more about creating Table blocks, see Sample 3. Step 3.

You have specified the properties of your FlexiLayout and can now create the required elements.

Step 4. Analyzing images to determine the order in which the elements should be detected

At this stage, you have to establish the following:

Is there any pattern or method in the arrangement of the fields on the images?

Which elements can be relied on to look for the data fields?

In what order should we look for the elements? (This is important, because at each subsequent step we can only rely on the elements of the previous step.)

As we are dealing with a multi–page document, we must first establish which objects can be used to identify the first and last pages of the document. These objects can be described by means of special compound **Header** and **Footer** elements.

The **Header** element must match only the first page of the document.
If a project also contains documents of other types, this elements will also be used as an identifier (i.e. a unique characteristic identifying this document type).

The **Footer** element must match the last page of the document only. We recommend creating required subelements in this group to prevent the **Footer** element from being matched with any other document page.

Once you have analyzed the images, you will notice that:

1. On the first page, there is a group of fields that consists of **Invoice Number**, **Invoice Date**, **Delivery Address**. The title of the **Invoice Number** field always goes at the beginning of each document, whereas the **Invoice Date** and **Delivery Address** are not always present.

   The **Invoice Number** and **Invoice Date** fields can be located either to the right of their corresponding titles or below them.

   For **Delivery Address**, we should also look either to the right or below the corresponding title, having limited the search area. Additionally, we will need an element to restrict the search area from below.

   Since on some of the images these fields have no values, you can speed up the matching process by specifying the following condition: do not look for the value of a field if the field's title have not been detected.
2. We can use this group of fields as an identifier for our document. We will describe these fields as part of a compound **Header** element named **InvoiceHeader**.
3. The last page of the document has the words **TOTAL AMOUNT MUST**, **Carried over**, **Total CHF**, **TOTAL** below the table. However, these words may also occur elsewhere on the document (e.g. in the title or in the body of the table). To find these words, we will need to use additional reference elements (e.g. table column titles). These reference elements will help us restrict the search area.
4. The elements describing the last page of the document will be part of a compound **Footer** element named **InvoiceFooter**.
5. For the **Footer** element to match only the last page of the document, it must contain a required element. As the words identifying the last page (see 3 above) occur on each last page of each document, we will make the element that describes them a required element.
6. The table (name it **InvoiceTable**) starts on the first page and ends on the last one. Additionally, the table is always preceded by the column titles on the first page. To identify the end of the table (on the last page), we will use an auxiliary element (e.g. the required element from the **InvoiceFooter** group).

   Note: The multitude of all pages of a document is called a multi–page canvas. A multi–page canvas is formed by joining all the pages of a document, top–down, without any gaps, the left border of all the pages lying on the same axis, which goes through the point (0, 0). The order in which the pages are joined together is determined by the order of the pages in the batch, therefore, we can only specify the start of the table (its header on the first page) and end of the table (its footer on the last page). The program will search for the table in the entire document, i.e. on the entire multi–page canvas.
7. We will look for the name of the company in the **Company** field always on the first page and always in the first upper third of the page.
8. The title of the **Total Amount** field is always located on the last page, below the table. The value of the field is located either to the right of the title or below it.

Step 5. Document header and InvoiceHeader group

In the **InvoiceHeader** group, we will describe the elements that will help us identify the first page of the document.

1. To find the keywords contained in the titles of the **Invoice Number**, **Delivery Address**, and **Invoice Date** fields, create the following elements of the **Static Text** type:

    kwInvoiceNumber element. See **Step 5.1** for detailed instructions.

    kwDeliveryAddress element. See **Step 5.2** for detailed instructions.

    kwInvoiceDate element. See **Step 5.3** for detailed instructions.

2. We will search for the **Invoice Number** and **Invoice Date** fields to the right and below the corresponding title, using the following elements:

    InvoiceNumber of the **Character String** type. See **Step 5.4** for detailed instructions.

    InvoiceDate of the **Date** type (for good quality images) and **InvoiceDateAsString** of the **Character String** type (if the search for the **InvoiceDate** element produced no results) merged into a **grDate** group element. See **Step 5.5** for detailed instructions.

    Note: For more information about searching for dates on poor quality images, see Help>Tips and Tricks.

3. To find the **Delivery Address** field, which has multiple lines, use a **DeliveryAddress** of the **Text Fragment** type (see **Step 5.6** for detailed instructions). The search area should be specified as strictly as possible.
    We will use a **wgAddressAbove** element of the **White Gap** type to restrict the search area from below (see **Step 5.6** for detailed instructions).
    Next will group all these elements that serve to detect the **Delivery Address** field into one group named **grAddress**:

Step 5.1. title of Invoice Number field, kwInvoiceNumber element

Once you have examined the images, you will notice that:

*The title of the **Invoice Number** field is present on all the images and it is the title that can be used to differentiate them from other documents. For this reason , we will make the title of the **Invoice Number** field a required element, i.e. it will be used as an identifier for this document type. For details, see Help> Elements\Required and optional elements.*

*The title may contain one of the following phrases: **Invoice no.**, **Credit note No**, **Invoice Number:**, **Invoice Number**. We will list all the possible variants in the search string. Type each phrase without spaces to make the maximum number of errors to apply to the entire phrase and not to separate words.*

We are now interested only in the field located on the first page of each document, in the first part of the first page (some of the first pages contain static text which is similar to the title but located at the bottom of the page).

To create a **kwInvoiceNumber** element:

1. In the **InvoiceHeader** element, create an element of **Static Text** type and name it **kwInvoiceNumber**.
2. Clear the default **Optional element** option on the **General** tab (earlier we decided to make this element required).
3. Click the **Static Text**.
4. In the **Search Text** field, type the search text: **InvoiceNo.|CreditNoteNo|InvoiceNumber:|InvoiceNumber**
5. Set **Max error percentage** to 20.
6. Click the **Search Constraints** tab and use absolute coordinates to specify the rectangular search area on the first page:

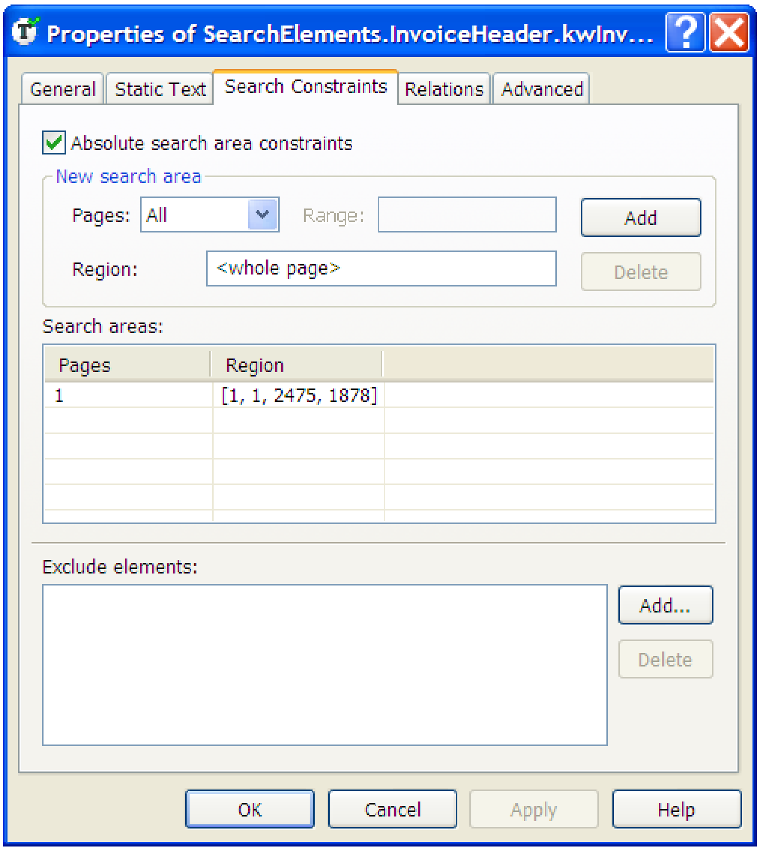
    Select **Absolute search area constraints**

    In the **Pages** field, select **Range**

    In the **Range** field, type 1

    In the **Search region** field, specify the coordinates of the rectangle [1, 1, 2475, 1878]

    Once you have specified the properties of the search area, click **Add**. The search area will appear in the **Search areas** list



☑ Absolute search area constraints

Pages: All

Range:

Region: <whole page>

Pages	Region
1	[1, 1, 2475, 1878]

7. Click the **Advanced**
8. In the **Advanced pre-search relations** field, specify additional constraints:

   WholeWordMode: true; // Only whole text objects will be treated as possible hypotheses.

   MaxGapInLine: 50dt; // The maximum distance between neighboring words.

9. By specifying an additional check in the **Advanced post-search relations** field, you can influence the quality of the generated hypotheses. For example: *If the element has been detected, calculate the difference between the reference width and the width of the region of the detected element and check if it belongs to the fuzzy interval.*
   In the FlexiLayout language, this condition can be written as follows:

   If not IsNull then

   { FuzzyQuality: 600dt – width, {-70000,0,0,70000}*dt;

   }

   //This condition means that the greater the difference, the greater the penalty/

   Note:. To specify the locations of objects and the distances among them, a two-dimensional coordinate system is used. The (0, 0) point, where the vertical and horizontal axes intersect, is located in the top left corner of the image. The abscissa axis is directed left to right, the ordinate axis is directed top to bottom.

10. Temporarily exclude the **InvoiceFooter** element and test the FlexiLayout.
    Note: To exclude an element, select it in the tree of elements and then select **Disable** from the **FlexiLayout** menu.

Step 5.2. title of Delivery Address field, kwDeliveryAddress element

Once you have examined the images, you will notice that:

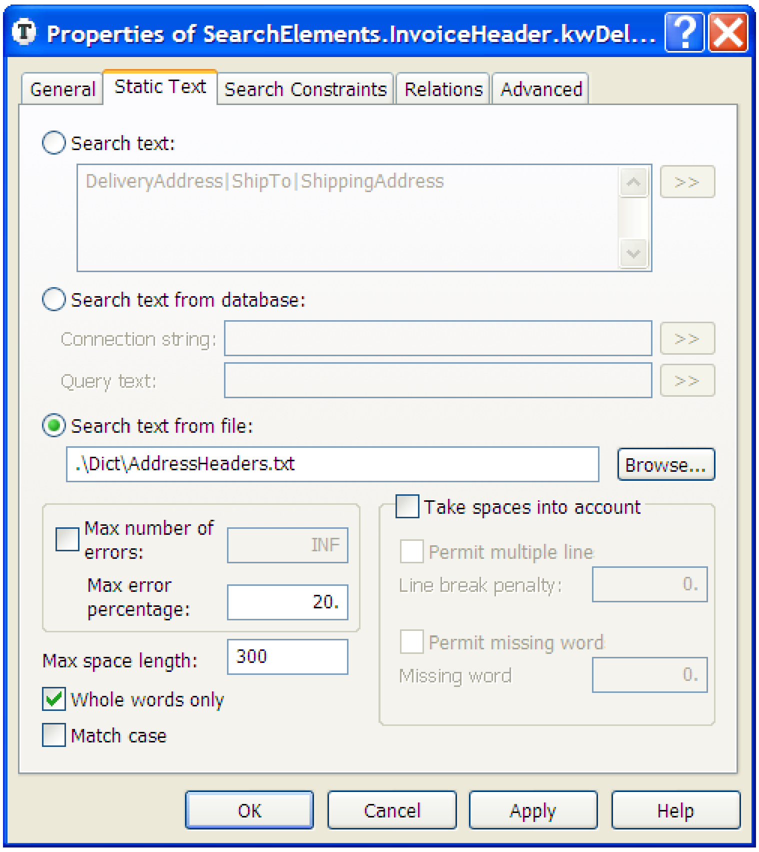
> *The field in question is always located on the first page, in the first part of the page. Therefore, you can limit the search area to the first page and use absolute coordinates to specify a rectangular area.*
>
> *We will use the **AddressHeaders.txt** file to search for the text.*
> ***Note**: Both the GUI and the FlexiLayout language allow you to use databases and text files to specify text lines to be found on an image. Text files can use either ANSI or Unicode encoding. Letter case does not affect the search and only serves to make the text more readable. Each search string must be placed on a separate line. Empty lines are ignored.*

You can set the maximum error percentage to 20, which in our case means no more than 3 errors in each phrase.
Note: *The optimal number of errors can only be arrived at through trial and error. For details about the* ***Static Text*** *elements, see Help>Elements\Elements Properties\ Static Text.*

To create a **kwDeliveryAddress** element:

1. In the **InvoiceHeader** element, create an element of **Static Text** type and name it **kwDeliveryAddress**.
2. Click the **Static Text** tab.
3. Select **Search Text from File** and provide the path to the **AddressHeaders.txt** file.

Note: The **AddressHeaders.txt** file can be found in **<disk name>:\\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\9.0\Samples\FlexiLayout Studio\Invoice\Dict**.
A text file is accessed by its name, which can be either absolute or relative. Relative names are specified relative to the FlexiCapture project folder (the folder that contains the *.fcproj file) or the FlexiCapture Studio project folder (the folder that contains the images).



Properties of SearchElements.InvoiceHeader.kwDel...

○ Search text:
DeliveryAddress|ShipTo|ShippingAddress

○ Search text from database:
Connection string:
Query text:

● Search text from file:
.\Dict\AddressHeaders.txt

□ Max number of errors: INF
Max error percentage: 20.
Max space length: 300
☑ Whole words only
□ Match case

□ Take spaces into account
□ Permit multiple line
Line break penalty: 0.
□ Permit missing word
Missing word 0.

4. Set **Max error percentage** to 20.
5. Click the **Search Constraints** tab and use absolute coordinates to specify the rectangular search area on the first page:

   Select **Absolute search area constraints**

   In the **Pages** field, select **Range** from the drop–down list

   In the **Range** field, type 1

   In the **Search region** field, specify the coordinates of the rectangle [1, 1, 2475, 1878]

   Once you have specified the properties of the search area, click **Add**. The search area will appear in the **Search areas** list

6. Click the **Advanced** tab and in the **Advanced pre–search relations** specify WholeWordMode: true;
// Only whole text objects will be treated as possible hypotheses.
7. Temporarily exclude the **InvoiceFooter** element and test the FlexiLayout.

Step 5.3. title of Invoice Date field, kwInvoiceDate element

Once you have examined the images, you will notice that:

> *The title of the **Invoice Date** field is a short and very common word **Date**.*
> *This means that we have to narrow down the search area as much as possible.*
>
> *The title of the **Date** field (if this field is present on a given image) is located very close to the title of the **Invoice Number** field.*

To create a **kwInvoiceDate** element:

1. In the **InvoiceHeader** element, create an element of **Static Text** type and name it **kwInvoiceDate**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, type the text to search: **Date**.
4. Set **Max error percentage** to 30 (which means that only one error can be made in our word which has four letters).
5. Click the **Relations** tab.
6. On the **Relations** tab, specify that the keywords may not be more than 30 dots above the upper border of the title of the **Invoice Number** field and not more than 700 dots below its lower border:

   > **Below** the upper border of the region of **kwInvoiceNumber**, **Offset** = –30, **Element border** = **Top**.
   >
   > **Above** the lower border of the region of **kwInvoiceNumber**, **Offset** = –700, **Element border** = **Bottom**.

   Note: Once you have created the new relation, click **Add** The new relation will appear in the **Relations**.

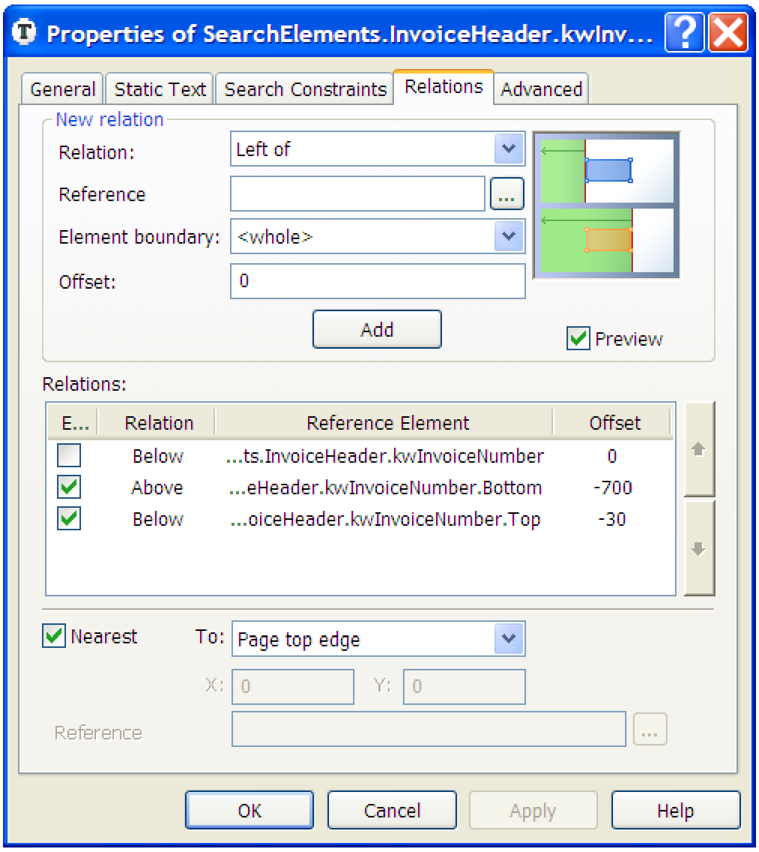
   Note: When selecting **Offset** values, you may wish to examine the geometrical properties (borders and sizes) of the reference and search image objects in pre–recognition results mode. You can use the **Measure Rectangle** tool to measure distances. **Offset** values can only be selected by method of trial and error.
7. Temporarily exclude the **InvoiceFooter** element and match the FlexiLayout.

Try matching the FlexiLayout. You will notice that:

> On the first page of some of the documents, several hypotheses have been generated for **kwInvoiceDate** and sometimes the wrong hypotheses has been selected. To correct the error, we need to adjust the properties of the element.
>
> The field we are interested in is closest to the upper border of the page.

To specify additional search constraints:

1. Open the **Properties** dialog box for the **kwInvoiceDate** element.
2. Click the **Relations** tab.
3. Select **Nearest** and then select **Page top edge** from the **To** drop–down list.



Relation: Left of

Reference

Element boundary: <whole>

Offset: 0

☑ Preview

E...	Relation	Reference Element	Offset
□	Below	...ts.InvoiceHeader.kwInvoiceNumber	0
☑	Above	...eHeader.kwInvoiceNumber.Bottom	-700
☑	Below	...oiceHeader.kwInvoiceNumber.Top	-30

☑ Nearest To: Page top edge

X: 0 Y: 0

Reference

Test the FlexiLayout again to make sure that the title of the **Invoice Date** field can be reliably detected.

Step 5.4. Invoice Number field, InvoiceNumber element

On some of the documents, the ***Invoice Number*** *field is located to the right of the field title whereas on other documents it is located below the title (in documents 1 and 2 the field data are placed directly beneath the title).*

To create an **InvoiceNumber** element:

1. In the **InvoiceHeader** element, create an element of **Character String** type and name it **InvoiceNumber**.

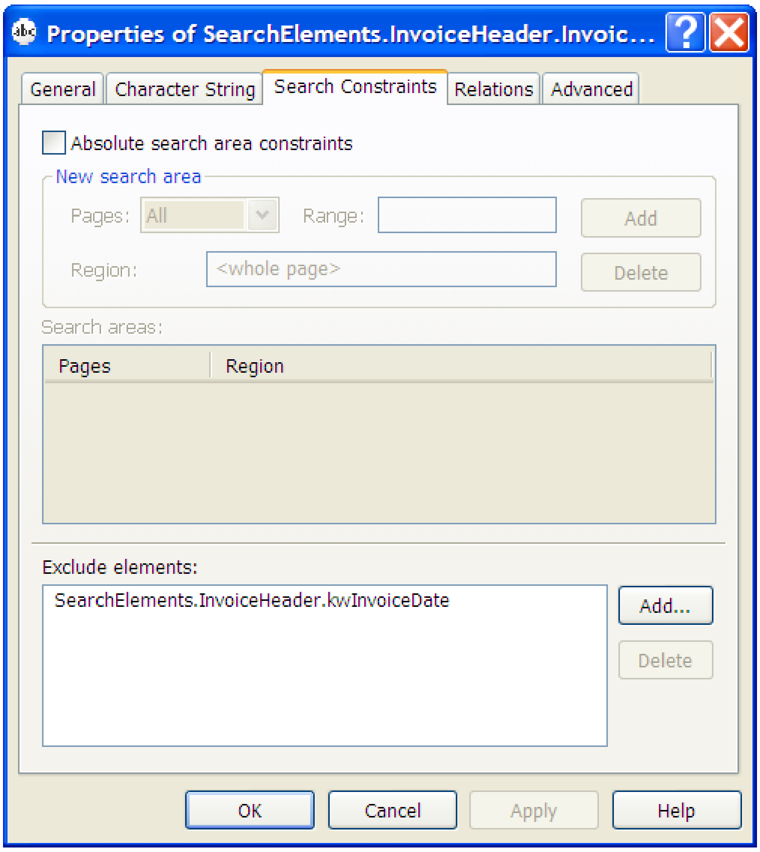
   Note: The name of the element need not be the same as the name of the block corresponding to the **InvoiceNumber** field. However, having identical names is convenient when working with a FlexiLayout.

2. Click the **Character String** tab.
3. Specify the alphabet:
   –./0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZabcdefghijklmnopqrstuvwxyz
4. Set **Percentage of non–alphabet characters** to 30%.
5. In the **String length** field, specify the fuzzy interval {–1, 3, 20, INF} for the length of the character string, assuming that possible values fall into the region of 3 to 20 characters. Any hypothesis outside this interval will be penalized.
6. On some of the images, the search area for the **Invoice Number** field contains the already detected **kwInvoiceDate** element. Exclude this element from the search area, so that the program does not include this text in its hypotheses for the **Invoice Number** field:

   Click the **Add...** button next to the **Exclude elements** field.

   Select **kwInvoiceDate** from the list of elements.

   Click **OK**. The string **SearchElements.InvoiceHeader.kwInvoiceDate** will appear in the **Exclude elements** field.



□ Absolute search area constraints

Region: <whole page>

Exclude elements:

SearchElements.InvoiceHeader.kwInvoiceDate

7. Click the **Advanced** tab.
8. In **Advanced pre-search relations**, specify additional conditions: if the title **kwInvoiceNumber** is not found, do not search for the **InvoiceNumber** field, otherwise specify the search area as an array of rectangles: one rectangle to the right of the title and another rectangle below the title with a small offset. In the FlexiLayout this condition can be written as follows:

```
if kwInvoiceNumber.IsNull then Dontfind();
else
{ WholeWordMOde: true;
MaxGapInLine: 30dt;

RectArray DataRegion;

Let r1= Rect (kwInvoiceNumber.Rect.Right, kwInvoiceNumber.Rect.Top –20dt, kwInvoiceNumber.Rect.Right + 650dt,
InvoiceNumber.Rect.Bottom + 50dt);
// Limits the search area by a rectangle to the right of the title of the InvoiceNumber field.

Let r2 = Rect (kwInvoiceNumber.Rect.Left – 100dt, kwInvoiceNumber.Rect.Bottom, kwInvoiceNumber.Rect.Right + 100dt,
wInvoiceNumber.Rect.Bottom + 100dt);
// Limits the search area by a rectangle beneath the title of the InvoiceNumber field.

DataRegion = RectArray (r1);
DataRegion.Add (r2);
RestrictSearchArea (DataRegion);
}
```

9. The search area for the field is limited to two rectangles, one to the right of the field title and one beneath it. If the field data are located to the right of the field title, there is often some other text beneath the field (on the test images, this is most commonly the title and data of the Date field). Since this extraneous text will be included into the search area for the **Invoice Number** field, the program will generate a hypothesis for this text. To influence the quality of this hypothesis, specify additional checks in the **Advanced post-search relations** field: *the lower the text beneath the field title, the bigger the penalty*. In the FlexiLayout language, this condition can be written as follows:

```
if not IsNull then
{ FuzzyQuality: Top – kwInvoiceNumber.Rect.Top, {–30000,0,0,10000}*dt;
}
// The greater the distance between the title and upper border of the field, the bigger the penalty for the corresponding
hypothesis. In this particular case, the hypothesis for the element whose upper border coincides with the upper border of the
title will have the highest quality.
```

10. Temporarily exclude the **InvoiceFooter** element and test the FlexiLayout.

11. Specify the location of the **InvoiceNumber** block as a rectangular region of the **InvoiceNumber** element enlarged by 5 dots vertically and horizontally. To do this, select **Region expression** and type the following expression:

```
Rect outputRect;
outputRect = InvoiceHeader.InvoiceNumber.Rect;
IsNull = InvoiceHeader.InvoiceNumber.IsNull;
// Treats the block region as detected if the InvoiceNumber element has been detected.

OutputRegion = outputRect;
OutputRegion.Inflate (5dt, 5dt);
```

Step 5.5. Invoice Date field, grDate, InvoiceDate, and InvoiceDateAsString elements

Once you have examined the images, you will notice that:

> *On some of the documents, the **Invoice Date** field is located to the right of the field title and on others it is located beneath the title. Therefore, we will limit the search area to rectangles to the right and beneath the title.*
>
> *We will use a **Date** element to search for the date. Additionally, we will specify the following condition: if the field title is not detected, do not search for the field data.*
>
> *Quite often, dates may be recognized unreliably. This may be due to scanning defects, invalid date formats, etc. Therefore, we will specify an additional **Character String** element in case the **Date** element fails to find any date.*

To specify the settings common to all the elements, we will create a group element.
To create a **grDate** group element:

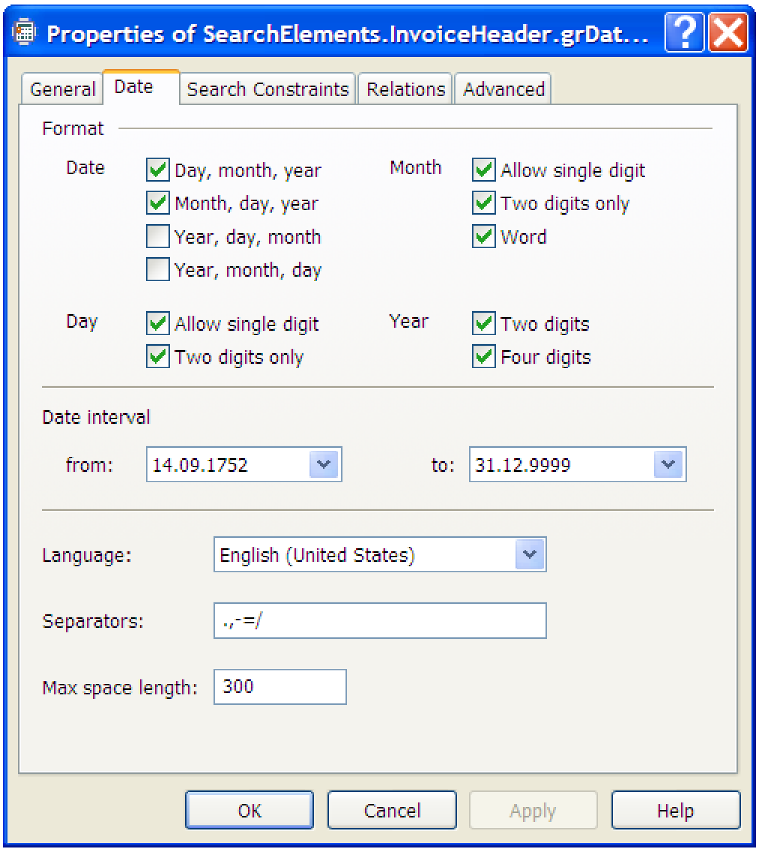
1. In the **InvoiceHeader** element, create an element of **Group** type and name it **grDate**.
2. Click the **Advanced** tab and specify additional search constraints: *limit the search area to the array of rectangles consisting of one rectangle to the right of the field title and rectangle below the field title with some offset*. In the FlexiLayout language this constraint can be written as follows:

```
RectArray DataRegion;
Let r1= Rect (kwInvoiceDate.Rect.Right, kwInvoiceDate.Rect.Top –20dt, kwInvoiceDate.Rect.Right + 650dt,
kwInvoiceDate.Rect.Bottom + 50dt);
Let r2 = Rect (kwInvoiceDate.Rect.Left – 150dt, kwInvoiceDate.Rect.Bottom, kwInvoiceDate.Rect.Right + 100dt,
kwInvoiceDate.Rect.Bottom + 100dt);

DataRegion = RectArray (r1);
DataRegion.Add (r2);
RestrictSearchArea (DataRegion);
```

To create an **InvoiceDate** element:

1. In the **InvoiceHeader.grDate** element, create an element of **Date** type and name it **InvoiceDate**.
2. Click the **Date** tab.
3. Specify all the possible date formats for the **InvoiceDate** element:



Format

Date
☑ Day, month, year
☑ Month, day, year
□ Year, day, month
□ Year, month, day

Month
☑ Allow single digit
☑ Two digits only
☑ Word

Day
☑ Allow single digit
☑ Two digits only

Year
☑ Two digits
☑ Four digits

Date interval

from: 14.09.1752 to: 31.12.9999

Language: English (United States)

Separators: .,-=/

Max space length: 300

4. On some of the images, the search area of the **Invoice Date** field will include the already detected **kwInvoiceNumber** and **InvoiceNumber** elements. To prevent the program from considering the values of these elements as hypotheses for the **Invoice Date** field, exclude these elements from the search area:

   Click the **Add...** button next to the **Exclude elements** field.

   Select **kwInvoiceNumber** from the list of elements.

   Click **OK**. The string **SearchElements.InvoiceHeader.kwInvoiceNumber** will appear in the **Exclude elements** field.

   Repeat the above actions for the element **SearchElements.InvoiceHeader.InvoiceNumber**.

5. Click the **Advanced** tab.

6. The **Invoice Date** field is not a required element. However, if a document contains a date (in the **Invoice Date** field), there is always the corresponding field title on the document (described earlier by the **kwInvoiceDate** element). Therefore, you can specify an additional search condition in **Advanced pre-search relations**: *Search for the image object only if the **kwInvoiceDate** has been detected.* In the FlexiLayout language, this condition can be written as follows:

   ```
   If InvoiceHeader.kwInvoiceDate.IsNull
   Then DontFind();
   ```

7. Temporarily exclude the **InvoiceFooter** element and test the FlexiLayout.

For poor quality images, when recognition results do not fit any of the standard parameters of the **Date** element, we will add an alternative element which will use more lax conditions to search for the **Invoice Date** field.
To create an **InvoiceDateAsString** element:

1. In the **InvoiceHeader.grDate** element, create an element of **Character String** type and name it **InvoiceDateAsString**.
2. Click the **Character String** tab.
3. Specify the alphabet:
   ,–./0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZ abcdefghijklmnopqrstuvwxyz
4. Set the **percentage of non-alphabet characters** to 30%.
5. In the **String length** field, specify the fuzzy interval {–1, 8, 14, INF}for the length of the character string, assuming that possible values fall into the region of 8 to 14 characters. Any hypothesis outside this interval will be penalized.
6. Set **Max space length** to 20. This limits the maximum length of space in the text string to 20 dots.

7. For the other element properties, keep the default settings.
8. Click the **Advanced** tab.
9. Since we are going to search for the **InvoiceDateAsString** element only if the **InvoiceDate** is not detected on the image, specify the following condition in the **Advanced pre–search relations** field: *Search for the image object only if the* ***InvoiceDate*** element is not detected. In the FlexiLayout language, this condition can be written as follows:

```
If Not InvoiceDate.IsNull Then DontFind;
```

10. Specify an additional condition for **InvoiceDate** similar to the one above: Search for the image object only if the **kwInvoiceDate** element has been detected. Search for an image object closest to the **kwInvoiceDate** element. In the FlexiLayout language, this condition can be written as follows:

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

11. Specify the location of the **InvoiceDate** block as the rectangular region of the detected **InvoiceDate** or **InvoiceDateAsString** element increased by 5 dots vertically and horizontally. To do this, select the **Region expression** option and type the following expression:

```
Rect outputRect;
if not InvoiceHeader.grDate.InvoiceDate.IsNull then
outputRect = InvoiceHeader.grDate.InvoiceDate.Rect;
else
{ outputRect = InvoiceHeader.grDate.InvoiceDateAsString.Rect;
IsNull = InvoiceHeader.grDate.InvoiceDateAsString.IsNull;
}
OutputRegion = outputRect;
OutputRegion.Inflate (5dt, 5dt);
```

Step 5.6. Delivery Address field, grAddress, wgAddressAbove, and DeliveryAddress elements

We have decided to search for the ***Delivery Address*** *field:*

to the right of and below the field title

using the ***White Gap*** *element to limit the search area from below*

using a ***Text Fragment*** *element*

To specify the properties common to all these elements, we will create a group element.

To create a group **grAddress** element:

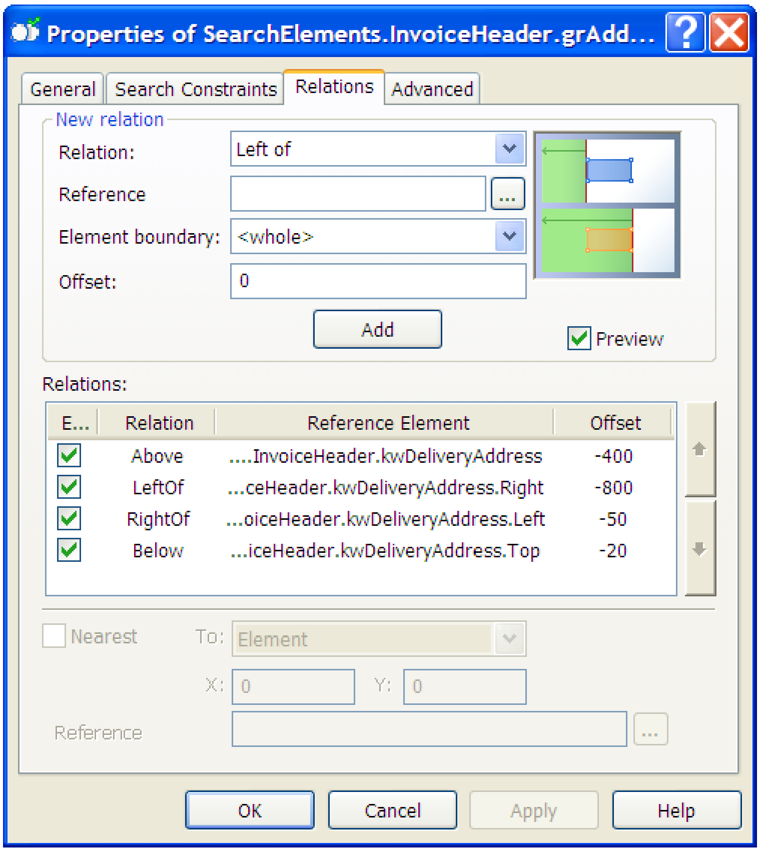
1. In the **InvoiceHeader** element create an element of **Group** type and name it **grAddress**.
2. Click the **Relations** tab and specify the following search constraints for the elements of the group:

Above the **kwDeliveryAddress** element, **Offset** = –400

Left of the right border of the **kwDeliveryAddress** element, **Offset** = –800, **Element border** = **Right**.

Right of the left border of the **kwDeliveryAddress** element, **Offset** = –50, **Element border** = **Left**.

Below the upper border of the **kwDeliveryAddress** element, **Offset** = –20, **Element border** = **Top**.



Properties of SearchElements.InvoiceHeader.grAdd...

Relation: Left of

Reference

Element boundary: <whole>

Offset: 0

☑ Preview

Relations:

E...	Relation	Reference Element	Offset
☑	Above	....InvoiceHeader.kwDeliveryAddress	-400
☑	LeftOf	...ceHeader.kwDeliveryAddress.Right	-800
☑	RightOf	...oiceHeader.kwDeliveryAddress.Left	-50
☑	Below	...iceHeader.kwDeliveryAddress.Top	-20

□ Nearest To: Element

X: 0 Y: 0

Reference

To limit the search area for the **DeliveryAddress** element from below, we will specify a horizontal **White Gap**.

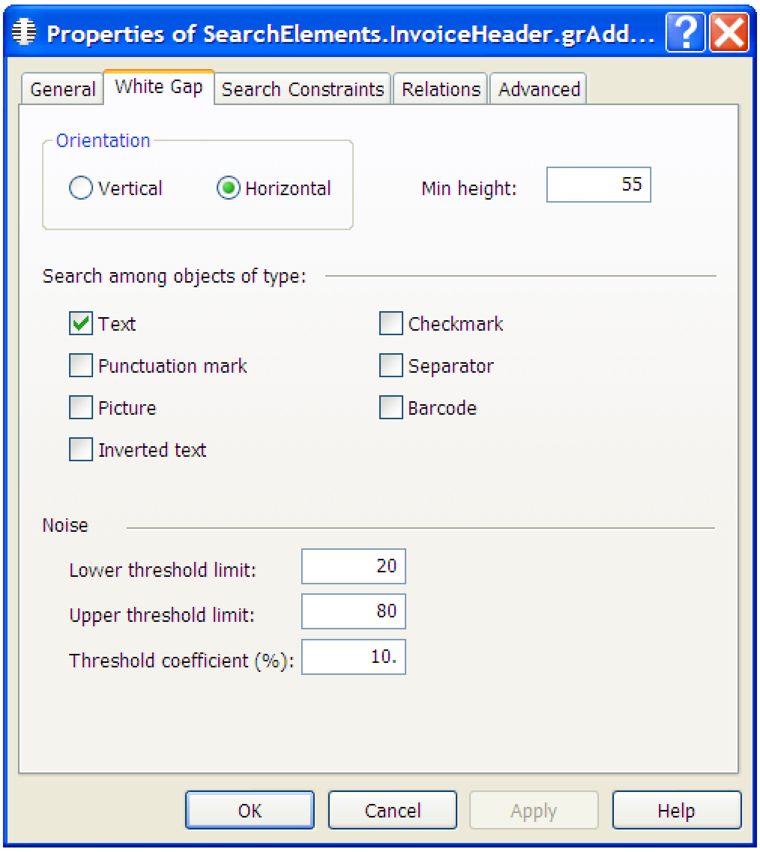
To create a **wgAddressAbove** element:

1. In the **InvoiceHeader.grAddress** element, create an element of the **White Gap** type and name it **wgAddressAbove**.
2. Click the **White Gap** tab.
3. Under **Orientation**, select **Horizontal**. Set **Min height** to 55.
   Note: When selecting values for the heights/widths of white gaps, you may wish to consider the geometry (i.e. borders and sizes) of the adjacent image objects in pre–recognition results mode.
4. We are going to look for a white gap inside a text fragment, therefore under **Search among objects of type**, select only **Text**.
5. Under **Rubbish**, specify the following values:

   Lower threshold limit – 20;

   Upper threshold limit – 80;

   Threshold coefficient (%) – 10.



Orientation

○ Vertical ● Horizontal

Min height: 55

Search among objects of type:

☑ Text ☐ Checkmark

☐ Punctuation mark ☐ Separator

☐ Picture ☐ Barcode

☐ Inverted text

Noise

Lower threshold limit: 20

Upper threshold limit: 80

Threshold coefficient (%): 10.

Note: The values for the properties **Lower threshold limit, Upper threshold limit** and **Threshold coefficient (%)** can only be selected by method of trial and error. We recommend starting with the default values. If the default values do not work, altering them until the white gap is detected on all the test images.

6. Do not specify any additional search constraints and retain the default values for the other options.
7. Click the **Relations** tab.
8. Since the white gap can be located only beneath the keywords of the **kwDeliveryAddress** title, specify the following search constraints:

   Below the **kwDeliveryAddress** element, **Offset** = 20

   Offset is required in order to prevent the program from finding white gaps in other areas on the image, e.g. between the field title and the text. Click the **Advanced** tab.
9. Since the **Delivery Address** field is not present on all of the images, but when it does occur it is always accompanied with a title, specify the following additional condition in the **Advanced pre-search relations** field: *Search for the image object only if the* ***kwDeliveryAddress*** element is detected. In the FlexiLayout language, this condition can be written as follows:

   If InvoiceHeader.kwDeliveryAddress.IsNull Then DontFind;
10. Test the FlexiLayout to see if the **wgAddressAbove** element is detected on all images.
11. Since the program sometimes finds more than one white gap in the specified search area and cannot always select the right one, we need to impose additional search constraints. If you examine the images, you will notice that the white gap we are interested in is the topmost gap in the given search area. Therefore, in the **Advanced pre-search relations** field, specify the following additional condition: *Search for an image object closest to the upper border of the first page*. In the FlexiLayout language, this condition can be written as follows:

    NearestY: Page(1).RectGlobal.Top;
12. Temporarily exclude the **InvoiceFooter** element and test the FlexiLayout.

To create a **DeliveryAddress** element:

1. In the **InvoiceHeader.grAddress** element, create an element of the **Text Fragment** type and name it **DeliveryAddress**.

2. Click the **Relations** tab.
3. Since **DeliveryAddress** can be located only above the white gap **wgAddressAbove**, specify the following search constraint:

    Above the **wgAddressAbove** element, **Offset** = 0.

5. Since the search area for the **DeliveryAddress** element has already been specified at the level of the **InvoiceHeader.grAddress** group, we need not specify it once again here.
6. Click the **Advanced** tab.
7. Since the title of the **DeliveryAddress** field is also located inside the search area, we need to exclude it in order to obtain only the value of **DeliveryAddress**. Additionally, there are separators and noise in the search area which are also best excluded. For the above reasons, we cannot just exclude the region of the **kwDeliveryAddress** element. It would be more appropriate to take the region of the **kwDeliveryAddress** element, enlarge this region so that it encompasses all the noise, and then exclude the resulting rectangle. To do this, specify the following search constraints in the **Advanced pre–search relations** field:

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

// Draws a rectangle around the field title (with some offset for the borders)
{ Let Left = InvoiceHeader.kwDeliveryAddress.Rect.Left – 100dt;
Let Top = InvoiceHeader.kwDeliveryAddress.Rect.Top – 100dt;
Let Right = InvoiceHeader.kwDeliveryAddress.Rect.Right + 20dt;
Let Bottom = InvoiceHeader.kwDeliveryAddress.Rect.Bottom;

// Exclude the resulting rectangle from the search area of the field
ExcludeRect (Left, Top, Right, Bottom);
}
```

8. Temporarily exclude the **InvoiceFooter** element and test the FlexiLayout.
9. Specify the location of the **DeliveryAddress** block as the rectangular region of the detected **DeliveryAddress** element, simplified for better visual perception and expanded by 5 dots vertically and horizontally. To do this, select the **Region expression** option and type the following expression:

```
Rect outputRect;
OutputRegion = InvoiceHeader.grAddress.DeliveryAddress.SimplifiedRegion;
IsNull = InvoiceHeader.grAddress.DeliveryAddress.IsNull;
OutputRegion.Inflate (5dt, 5dt);
```

Step 6. Document Footer, InvoiceFooter group

In the **InvoiceFooter** group we will specify the elements that will enable the program to identify the last page of the document:

> The last page of the document always has the keywords "**TOTALAMOUNTMUST**", "**Carriedover,**" "**TotalCHF,**" "**Total.**"
>
> To search for this text, we will create two elements. In the first element, specify "**TOTALAMOUNTMUST,**" "**Carriedover,**" and "**TotalCHF**" Create a second element below the first one in the tree of elements and specify only "**Total**" in this second element (the program will search for this element only if the first element has not been detected). This will reduce the number of hypotheses generated for the element and speed up the FlexiLayout matching process.
>
> These words are always located below the column titles **Description** and **Product Total** on the same page.

1. To search for the keywords in the **Product Total** and **Description** column, we will create elements of the **Static Text** type:

   ExtraTag element: see detailed instructions in **Step 6.1**.

   AddTag element: see detailed instructions in **Step 6.2**.

2. We will also use **Static Text** elements to search for the keywords "**TOTALAMOUNTMUST,**" "**Carriedover,**" "**TotalCHF,**" and "**Total**" :

   LongFooter element: see detailed instructions in **Step 6.3**.

   ShortFooter element (to search for the keyword "**Total**"): see detailed instructions in **Step 6.4**.

3. To mate **Footer** match only the last page of the document, it must contain a required element. Since the keywords that identify the last page of the document can be found on every last page of every document, in the **InvoiceFooter** group, create a required element of the **Text Fragment** type and name it **InvoiceFooter**. Specify the search area for this element as the rectangular region of the detected **LongFooter** or **ShortFooter** element, slightly expanded vertically and horizontally (for detailed instructions, see **Step 6.5**).

Step 6.1. Product Total column title, ExtraTag element

To create an **ExtraTag** element:

1. In the **InvoiceFooter** element, create an optional element of the **Static Text** type and name it **ExtraTag**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, type the text to search: **PRODUCTTOTAL**.
4. Do not impose any additional constraints and leave default values for the other options.

Step 6.2. Description column title, AddTag element

To create an **AddTag** element:

1. In the **InvoiceFooter** element, create an optional element of the **Static Text** type and name it **AddTag**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, type the text to search: **Designation|Description**.
4. In **Advanced pre-search relations** specify additional constraints:

```
WholeWordMode: true; // Only whole words must be considered as hypotheses.
MaxGapInLine: 10dt; // This is the maximum distance between adjacent characters.
```

Step 6.3. The final part of the document, LongFooter element

To create a **LongFooter** element:

1. In the **InvoiceFooter** element, create an optional element of the **Static Text** type and name it **LongFooter**.

2. Click the **Static Text** tab.
3. In the **Search Text** field, type the text to search: **TOTALAMOUNTMUST|Carriedover|TotalCHF**.
4. Since the search text is always located below the title of the **Description** column, click the **Relations** tab and specify the following relation:

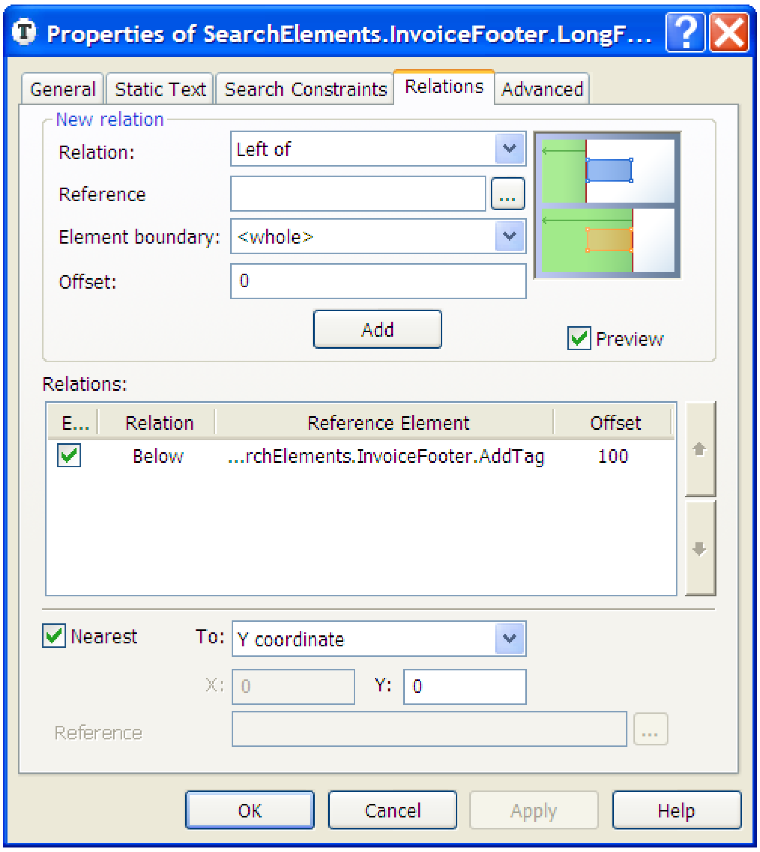
   Below the **AddTag** element, **Offset** = 100;

5. Since on some of the documents (e.g. on documents 3 and 9) the program finds several text variants and selects the wrong one, we must impose additional constraints. Once you have analyzed the images, you will notice that the text we are interested in is the topmost text in the given search area (we want to find the text that is closest to the table order, because it can then be used as the **Footer** of the table). On the **Relations** tab, specify additional search constraints: Search for a string closest to the upper border of the document:

   Select the **Nearest**

   In the **To** field, select **Y coordinate** from the drop–down list.

   In the **Y** field, type 0.


Properties of SearchElements.InvoiceFooter.LongF...

New relation

Relation: Left of

Reference:

Element boundary: <whole>

Offset: 0

☑ Preview

Relations:

E...	Relation	Reference Element	Offset
☑	Below	...rchElements.InvoiceFooter.AddTag	100

☑ Nearest To: Y coordinate

X: 0 Y: 0

Reference

Step 6.4. The final part of the document, ShortFooter element

To create a **ShortFooter** element:

1. In the **InvoiceFooter** element, create an optional element of the **Static Text** type and name it **ShortFooter**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, type the text to search: **Total**.
4. Click the **Search Constraints** tab and exclude the already detected **ExtraTag** element from the search area of the **ShortFooter** element. This will prevent the program from finding the word "**Total**" in the title of the table:

   Click the **Add...** button next to the **Exclude elements** field.

   Select **ExtraTag** from the list of elements.

   Click **OK**. The **SearchElements.InvoiceFooter.ExtraTag** string will appear in the **Exclude elements** field.

5. Since the search text is always located below the title of the **Description** column, click the **Relations** tab and specify the following relation:

   Below the **AddTag** element, **Offset** = 100;

5. Earlier we decided that we will search for the **ShortFooter** element only if **LongFooter** has not been detected. To implement this condition, specify the following in **Advanced pre–search relations**: *Do not search for the element if **LongFooter** has not been detected*. In the FlexiLayout language, this condition can be written as follows:

   ```
   If not LongFooter.IsNull then Dontfind();
   ```

Step 6.5. The final part of the document, TextFragment element

To create a **TextFragment** element:

1. In the **InvoiceFooter** element, create a required element of the **Text Fragment** type and name it **Text Fragment**.
2. Click the **Advanced** tab and in the **Advanced pre–search relations** field, specify additional search conditions: *Do not search for the element if neither, **LongFooter** nor **ShortFooter** has been detected. If **LongFooter** has been detected, specify the search area as the rectangular region of this element expanded by 10 dots vertically and horizontally. If **ShortFooter** has been detected, specify the search area as the rectangular region of this element expanded by 10 dots vertically and horizontally*. In the FlexiLayout language, this condition can be written as follows:

   ```
   If LongFooter.IsNull and ShortFooter.IsNull then Dontfind();
   // Do not search for the element if none of the elements are detected
   else
   { if not LongFooter.IsNull then
   // If LongFooter is detected, the search area is defined based on its region
   { LeftOf: LongFooter.Rect.Right + 10dt;
   RightOf: LongFooter.Rect.Left – 10dt;
   Below: LongFooter.Rect.Top – 10dt;
   Above: LongFooter.Rect.Bottom + 10dt;
   }
   else if not ShortFooter.IsNull then
   // If ShortFooter is detected, the search area is defined as the region of ShortFooter
   { LeftOf: ShortFooter.Rect.Right + 10dt;
   RightOf: ShortFooter.Rect.Left – 10dt;
   Below: ShortFooter.Rect.Top – 10dt;
   Above: ShortFooter.Rect.Bottom + 10dt;
   }
   }
   ```

Try matching the FlexiLayout and make sure that the pages are correctly separated into documents as defined by the **Header** and **Footer** elements. Once the FlexiLayout is matched with all the pages, the documents should be assembled as specified in the reference assembly created when we started work on this FlexiLayout.

Note: To apply the FlexiLayout to all test images, use keyboard shortcuts:

Activate the **Batch** window

Press **Ctrl + A** to select all the pages.

Press **Ctrl + Shift + E** to analyze all the selected pages.

Note: In the case of a multi–page document, be sure to apply the FlexiLayout to all the pages of the document, not just to some of the pages. Otherwise, the program may fail to apply the FlexiLayout to some of the documents.

Step 7. Table column titles, TableHeader group

We will use a **Table** element to search for the **Invoice Table** field. However, some preparations have to be done: we have to detect the column titles first. In this sample, auxiliary elements are used to search for column titles.

Therefore, before we start searching for the fields, we need to create a number of elements.

To described the column titles, create a **TableHeader** group, which will include the descriptions of all the column titles. Also, limit the search area to the first page.

To create a **TableHeader** element:

1. In the tree of elements, create an element of the **Group** type and name it **TableHeader**.
2. Click the **Search Constraints** tab and specify the search area for the element:

    Select the **Absolute search area constraints** option.

    In the **Pages** field, select **Range** from the drop–down list.

    In the **Range** field, type 1.

    In the **Search region** field, keep the default value **<whole page>**.

    Click the **Add** button.

We will use **Static Text** elements to search for the column titles.
In the **TableHeader** group, create the following elements:

1. **kwDesignation** element: see detailed instructions in **Step 7.1**,
2. **ExtraQtyTag** element: see detailed instructions in **Step 7.2**,
3. **kwQuantity** element: see detailed instructions in **Step 7.3**,
4. **kwUnitPrice** element: see detailed instructions in **Step 7.4**,
5. **kwTotal** element: see detailed instructions in **Step 7.5**,
6. **kwReference** element: see detailed instructions in **Step 7.6**,
7. **kwSales** element: see detailed instructions in **Step 7.7**,
8. **kwUnit** element: see detailed instructions in **Step 7.8**.

Step 7.1. Title of Designation column, kwDesignation element

*Once you have analyzed the pre–recognition results, you will notice that there are only two variants of the title – **Designation** and **Description**. Since we specified the search area at the level of the group, we need not specify it here.*

To create a **kwDesignation** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **kwDesignation**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: **Designation|Description**.
4. Keep the default values for the other options.
5. Match the FlexiLayout to make sure that the element is detected on the mages.

Step 7.2. Title of ExtraQuantity column, ExtraQtyTag element

Now we need to detected the **Quantity** title, which appears on the documents as "**Qty,**" "**Quantity,**" "**ORDERED,**" or "**QtyNet**". However, there is one more title in the table which may also contain the word "**Quantity**". This title also contains unique text. Therefore, we will first described the **ExtraQtyTag** and then exclude it from the search area of the **Quantity** title.

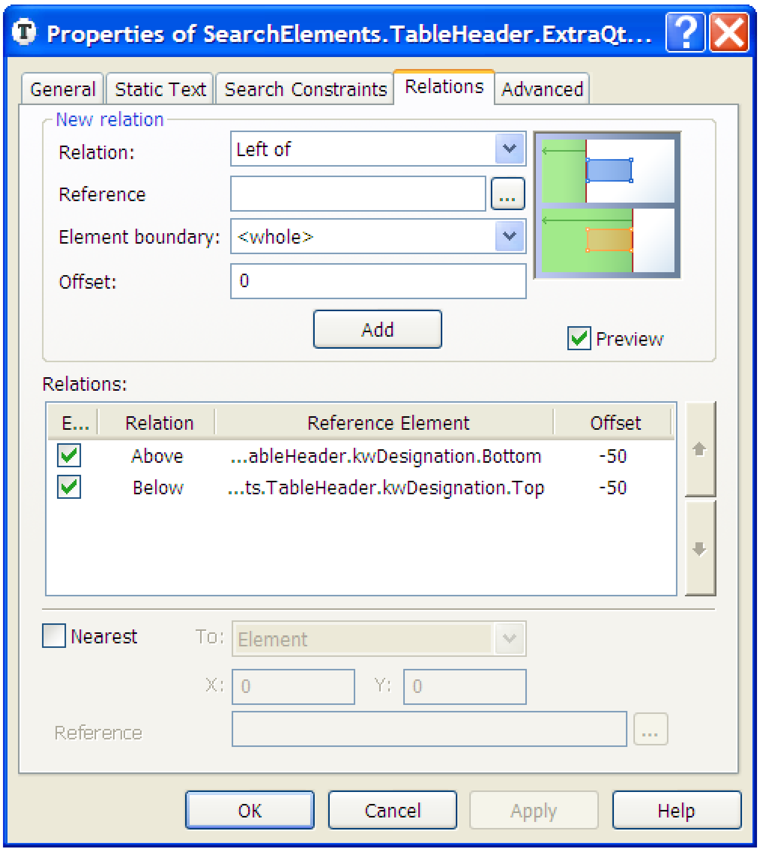
On all the documents, the title of the **ExtraQtyTag** column is exactly the same. Even though the quality of the images is fairly good, set the maximum error percentage to 35%. For a seven– or eight–letter word this means that two characters may be incorrect.

Additionally, since all the titles are placed on the same level, specify an additional search constraint relative to the already detected title of the **Designation** column.

To create an **ExtraQtyTag** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **ExtraQtyTag**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: {**Quantity** *__BK.ORD__}.
4. Set **Max error percentage** to 35.
5. Since the **ExtraQuantity** field contains multiple lines, select the options **Allow for spaces** and **Permit multiple lines**.
6. We will not penalize hypotheses carried over to the next line: set **Line break penalty** to 1.
7. To specify that the title we are searching is located on the same level with the title **Designation**, click the **Relations** tab and specify the following additional constraints:

   Above the lower border of the **kwDesignation** element, **Offset** = –50, **Element border** = **Bottom**,
   Below the upper border of the **kwDesignation** element, **Offset** = –50, **Element border** = **Top**.



Relation: Left of

Reference

Element boundary: <whole>

Offset: 0

☑ Preview

Relations:

E...	Relation	Reference Element	Offset
☑	Above	...ableHeader.kwDesignation.Bottom	-50
☑	Below	...ts.TableHeader.kwDesignation.Top	-50

□ Nearest To: Element

X: 0 Y: 0

Reference

8. To optimize the search process, specify an additional condition: *Do not search for the title if the title **Designation** has not been detected*. Click the **Advanced** tab and specify this condition in **Advanced pre-search relations**. In the FlexiLayout language, this condition can be written as follows:

   If kwDesignation.IsNull then Dontfind();

9. In **Advanced pre-search relations**, also specify additional constraints:

   WholeWordMode: true; // Only whole words will be considered as hypotheses.
   Note: This constrain can also be specified via the GUI by selecting **Whole words only** on the **Static Text** tab.

10. Match the FlexiLayout to make sure that the element is detected on the first page of the document.

Step 7.3. Title of Quantity column, kwQuantity element

We will rely on the already detected **kwDesignation** and **ExtraQtyTag** elements to describe this title.

To create a **kwQuantity** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **kwQuantity**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: **Qty|Quantity|ORDERED|QtyNet**.
4. Set **Max error percentage** to 35.
5. Click the **Search Constraints** tab and exclude the already detected element **SearchElements.TableHeader.ExtraQtyTag** from the search area.
6. To specify that the title we are searching is located on the same level with the title **Designation**, click the **Relations** tab and specify the following additional constraints. See **Step 7.2** for details.
7. To optimize the search process, specify an additional condition: *Do not search for the title if the title **Designation*** has not been detected. See **Step 7.2** for details.
8. In **Advanced pre-search relations**, specify additional search constraints:

```
WholeWordMode: true; // Only whole words will be considered as hypotheses.
MaxGapInLine: 20dt; // This is the maximum distance between adjacent characters.
```

9. Match the FlexiLayout to make sure that the element is detected on the first page of the document.

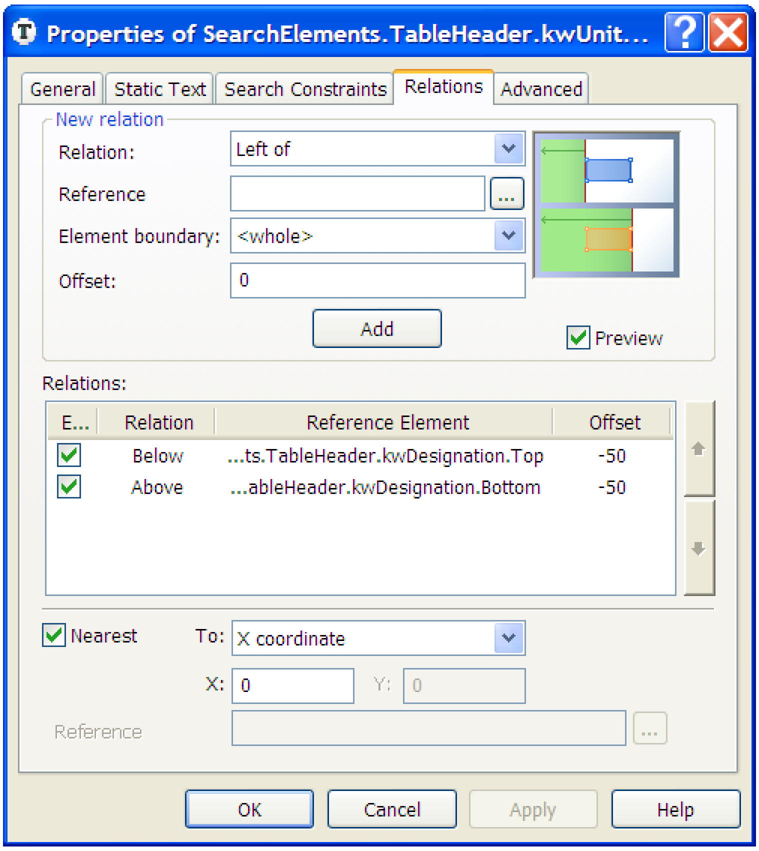
Step 7.4. Title of UnitPrice column, kwUnitPrice element

*Once you have analyzed the images you will notice that the title of the **Unit Price** column can be described by the following search text: "**UnitPrice,**" "**Price,**" and "**RATE**", and the title **Unit** appears as "**Unit.**" To prevent the program from detecting the word "**Unit**" in the title of the **Unit Price** column, we will first describe **Unit Price**, because it contains unique text, and then we will use an additional relation do describe **Unit**.*

*When creating an element to search for the keywords of the title of the **Unit Price** column, take into account the fact that the **Unit Price** column is always located to the right of the **Quantity** column.*

To create a **kwUnitPrice** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **kwUnitPrice**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: **UnitPrice|Price|RATE**.
4. To specify that the title we are searching is located on the same level with the title **Designation**, click the **Relations** tab and specify the following additional constraints. See **Step 7.2** for details.
5. On the **Relations** tab, specify the following condition: search for an element closest to the **X coordinate** with value = 0.


Properties of SearchElements.TableHeader.kwUnit...

Relation: Left of

Reference

Element boundary: <whole>

Offset: 0

☑ Preview

Relations:

E...	Relation	Reference Element	Offset
☑	Below	...ts.TableHeader.kwDesignation.Top	-50
☑	Above	...ableHeader.kwDesignation.Bottom	-50

☑ Nearest To: X coordinate

X: 0 Y: 0

Reference

6. Click the **Advanced** tab.
7. Since the title of the **Unit Price** column is located to the right of the title of the **kwQuantity** column, specify the following additional conditions in **Advanced pre-search relations**: *If the **kwQuantity** element has been detected, search for the image object to the right of the **kwQuantity** element.* In the FlexiLayout language, this condition can be written as follows:

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

8. To optimize the search process, specify an additional condition: *Do not search for the title if the title **Designation*** has not been detected. **See Step 7.2** for details.
9. We can influence the quality of generated hypotheses by specifying additional checks in the **Advanced post-search relations** field. In this case, we may specify the following check: *If the element has been detected, calculate the difference between the reference width and the width of the detected element and see if it belongs to the given fuzzy interval.* In the FlexiLayout language, this condition can be written as follows:

```
If not IsNull then
{ FuzzyQuality: 400dt – width, {-50000,0,0, 50000}*dt;
}
//This check means that the greater the difference, the greater the penalty coefficient
```

10. Match the FlexiLayout to make sure that the element is detected on the first page of the document.

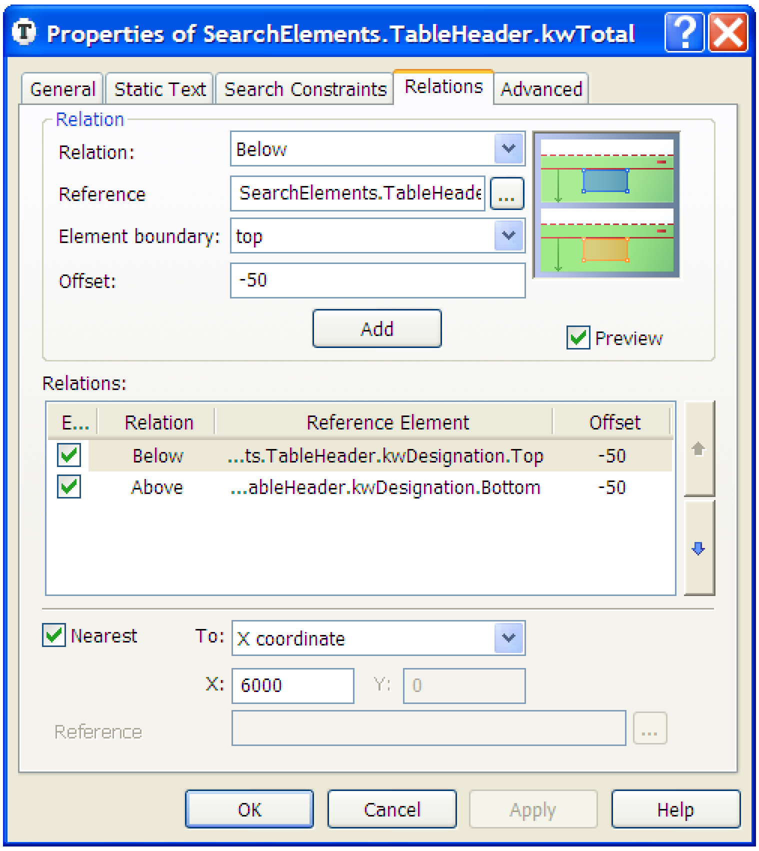
Step 7.5. Title of Total column, kwTotal element

You will notice that the **Total** column on all the images is located to the right of the **Unit Price** column. Therefore, we can search for the **Total** column relative to the **Unit Price** column.

To create a **kwTotal** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **kwTotal**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: **TotalChf|Total|Extension|AMOUNT|PRODUCTTOTAL**.

4. To specify that the title we are searching is located on the same level with the title **Designation**, click the **Relations** tab and specify the following additional constraints. See **Step 7.2** for details.
5. On the **Relations** tab, specify the following condition: search for an element closest to the **X coordinate** with value=6000.



Properties of SearchElements.TableHeader.kwTotal

Relation

Relation: Below

Reference SearchElements.TableHeade

Element boundary: top

Offset: -50

☑ Preview

Relations:

E...	Relation	Reference Element	Offset
☑	Below	...ts.TableHeader.kwDesignation.Top	-50
☑	Above	...ableHeader.kwDesignation.Bottom	-50

☑ Nearest To: X coordinate

X: 6000 Y: 0

Reference

6. Click the **Advanced** tab.
7. Since the title of the **Total** column is located to the right of the title of the **kwUnitPrice** column (if the latter is present on the image), specify an additional condition in **Advanced pre-search relations**: *If the element **kwUnitPrice** is detected, search for **kwUnitPrice***. In the FlexiLayout language, this condition can be written as follows:

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

8. To optimize the search process, specify an additional condition: *Do not search for the title if the title **Designation*** has not been detected. See **Step 7.2** for details.
9. Match the FlexiLayout to make sure that the element is detected on the first page of the document.

Step 7.6 Title of Reference column, kwReference element

You will notice that on the documents this title appears as "**Reference,**" "**ID,**" "**Item**" or "**CODE.**"

To create a **kwReference** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **kwReference**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: **Reference|ID|Item|CODE**.
4. To specify that the title we are searching is located on the same level with the title **Designation**, click the **Relations** tab and specify the following additional constraints. See **Step 7.2** for details.
5. To optimize the search process, specify an additional condition: *Do not search for the title if the title **Designation*** has not been detected. See **Step 7.2** for details.
6. Match the FlexiLayout to make sure that the element is detected on the first page of the document.

Step 7.7. Title of Sales column, kwSales element

Even we do not need the **Sales** column per se, it can be used to limit the table on the right.

To create a **kwSales** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **kwSales**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: **Sales**
4. To specify that the title we are searching is located on the same level with the title **Designation**, click the **Relations** tab and specify the following additional constraints. See **Step 7.2** for details.
5. Click the **Advanced** tab.
6. Since the title of the **Sales** column is located to the right of the **kwTotal** column (if the latter is present on the image), specify an additional condition in **Advanced pre-search relations**: *If the **kwTotal** element is detected, search for the image object to the right of the **kwTotal*** element. In the FlexiLayout language, this condition can be written as follows:

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

7. To optimize the search process, specify an additional condition: *Do not search for the title if the title **Designation*** has not been detected. See **Step 7.2** for details.
8. Match the FlexiLayout to make sure that the element is detected on the first page of the document.

Step 7.8. Title of Unit column, kwUnit element

When creating an element to search for the keywords of the title of the **Unit** column, we will rely on the already detected column **UnitPrice** and specify an additional relation: *Search for the title of the **Unit** column to the left of the left border of the title of the **UnitPrice** column.*

To create a **kwUnit** element:

1. In the **TableHeader** element, create an element of the **Static Text** type and name it **kwUnit**.
2. Click the **Static Text** tab.
3. In the **Search Text** field, specify the text to search: **Unit**.
4. To specify that the title we are searching is located on the same level with the title **Designation**, click the **Relations** tab and specify the following additional constraints. See **Step 7.2** for details.
5. To specify that the title is located to the left of the left border of the title of the **UnitPrice** column, specify the following additional constraint on the **Relations** tab:

   Left of the left border of the **kwUnitPrice** element **Offset** = 20, **Element border** = **Left**;

6. To optimize the search process, specify an additional condition: *Do not search for the title if the title **Designation*** has not been detected. See **Step 7.2** for details.
7. Match the FlexiLayout to make sure that the element is detected on the first page of the document.

Step 8. Table element, InvoiceTable element

Now that all the preparations have been made we can search for the **Invoice Table** field by creating a **Table** element:

1. **Header** of the table. As decided earlier, we will specify column titles using already detected auxiliary elements.
2. **Footer** of the table. To specify the **Footer**, we decided to use an already detected required element **TextFragment** from the **InvoiceFooter** group, which identifies the beginning of the final part. We described this element as the closest to the bottom of the table.
3. Search area for the table. We can limit the search area on the right using the title of the last **Sales** column, which we do not need for any other purposes.

To create an **InvoiceTable** element:

1. In the **SearchElements** tree, create and element of the **Table** type and name it **InvoiceTable**.
2. Click the **Columns** tab
3. Click "**...**" to select the previously created (see Step 3) **InvoiceTable** block as the block to search
4. Specify the search parameters for the table columns:

   4.1. In the list of columns, select **Reference** and click the **Properties**... button.
   For the title of this column, we will use the previously deselected **TableHeader.kwReference** element. To use it, select **Use found element as column title**, then click "**...**" and select **TableHeader.kwReference**.

   4.2. In the list of columns, select **Designation** and click the **Properties**... button.
   For the title of this column, we will use the previously deselected **TableHeader.kwDesignation** element.

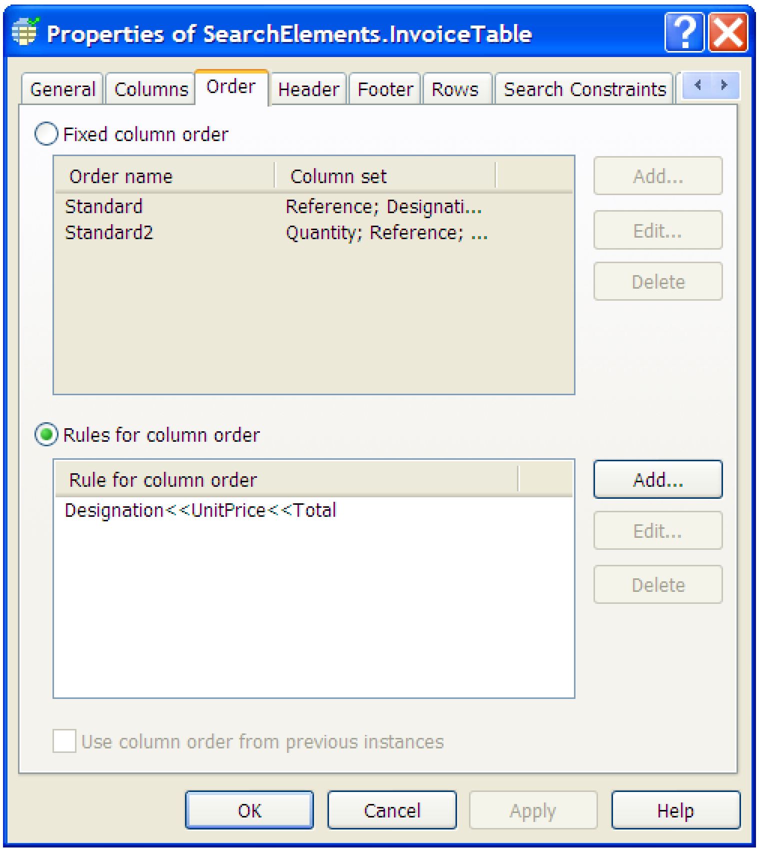
   4.3. In the list of columns, select **Unit** and click the **Properties**... button.
   For the title of this column, we will use the previously deselected **TableHeader.kwUnit** element.

   4.4. In the list of columns, select **Quantity** and click the **Properties**... button.
   For the title of this column, we will use the previously deselected **TableHeader.kwQuantity** element.

   4.5. In the list of columns, select **UnitPrice** and click the **Properties**... button.
   For the title of this column, we will use the previously deselected **TableHeader.kwUnitPrice** element.

   4.6. In the list of columns, select **Total** and click the **Properties**... button.
   For the title of this column, we will use the previously deselected **TableHeader.kwTotal** element.

5. Click the **Order** tab.
6. Since the columns we are looking for may be separated from one another by other columns (e.g. **ExtraQtyTag**), but their order is always fixed (first **Designation**, next **Unit Price**, next **Total**), select **Rules for column order**, click "**Add...**", and add the columns in the following order: **Designation** << **Unit Price** << **Total**.



Properties of SearchElements.InvoiceTable

○ Fixed column order

Order name	Column set
Standard	Reference; Designati...
Standard2	Quantity; Reference; ...

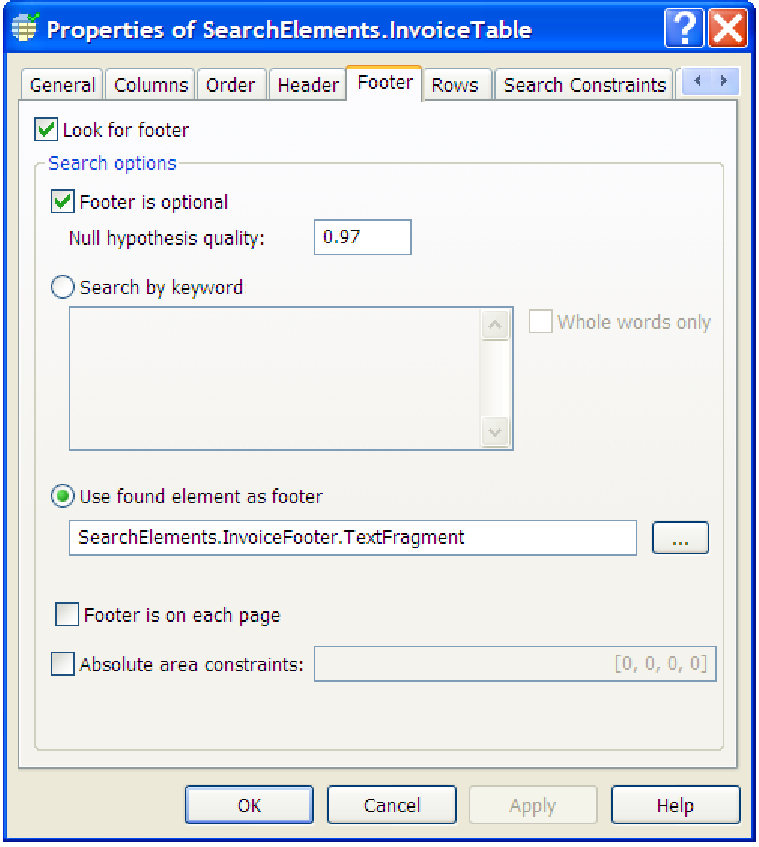
● Rules for column order

Rule for column order
Designation<<UnitPrice<<Total

□ Use column order from previous instances

7. Click the **Header** tab. We need the **Header** of the table, but we do not need to impose any additional constraints, so keep the default values.

8. Click the **Footer** tab. We will use the previously detected element **InvoiceFooter.TextFragment** for the final part of the table. Select the **Use found element as footer** option, click "**...**," and then select the **InvoiceFooter.TextFragment** element.



Properties of SearchElements.InvoiceTable

☑ Look for footer

Search options

☑ Footer is optional

Null hypothesis quality: 0.97

○ Search by keyword

□ Whole words only

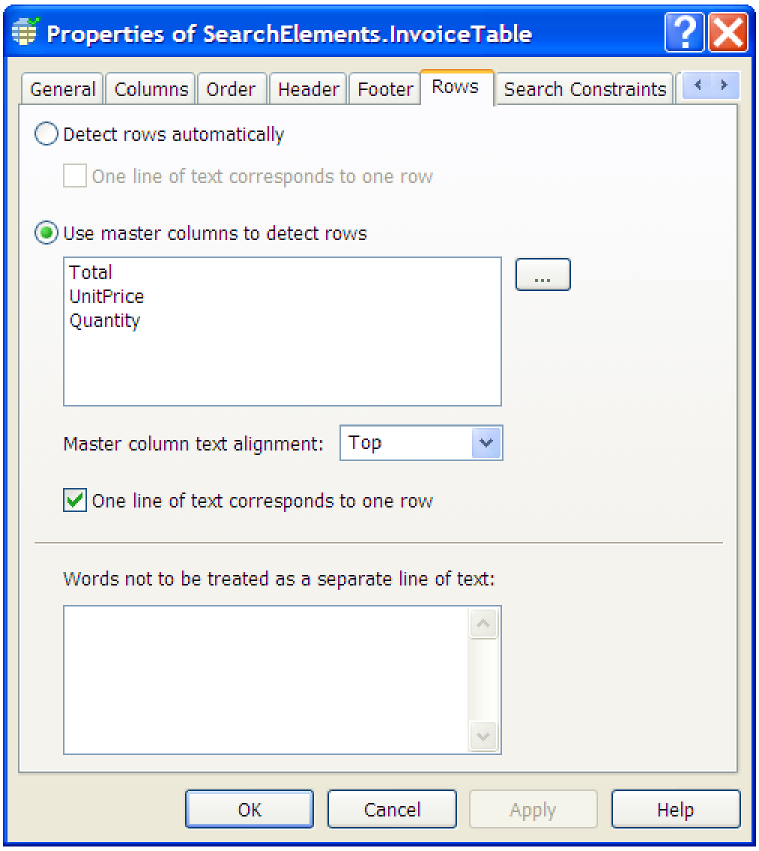
● Use found element as footer

SearchElements.InvoiceFooter.TextFragment

□ Footer is on each page

□ Absolute area constraints: [0, 0, 0, 0]

9. Click the **Rows** tab.
10. Since the cells in some of the columns may contain multiple lines, we need to specify master columns for line separation. Select **Use master columns to detect rows**.
    First specify the **Total** column, as this column always contains data and for each table record this column has one line of data. In case this column is no detected on the image, specify the **Unit Price** and **Quantity** columns as the next line separation option. Use the "**...**" button to select these columns.



○ Detect rows automatically

□ One line of text corresponds to one row

● Use master columns to detect rows

Total
UnitPrice
Quantity

Master column text alignment: Top

☑ One line of text corresponds to one row

Words not to be treated as a separate line of text:

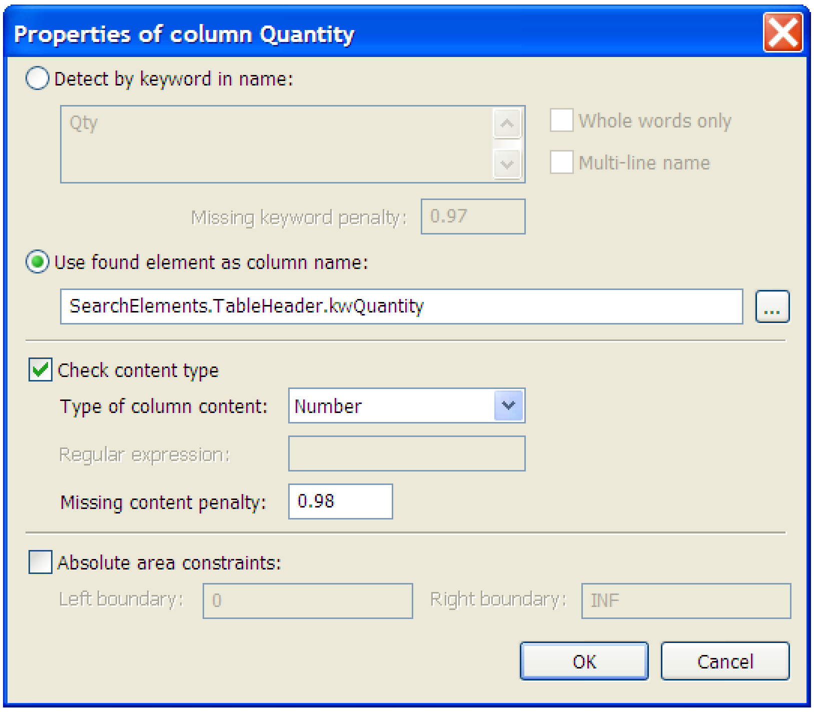
11. Select the **Relations** tab. Since the table columns we are interested in are located to the left of the title of the **Sales** column, add the following search constraints:

    Left of the **TableHeader.kwSales** element, **Offset** = 0;

    Note: For poor quality images (e.g. there is noise between the body and footer of the table, or there is no header or footer) you may whish to specify upper and lower table limits. In this case, we can use the **kwDesignation** element as an upper limit and the **TextFragment** element as the lower limit. On the **Relations** tab, add the following search constraints:

    Below the upper border of the **kwDesignation** element, **Offset** = –50, **Element border** = **Top**;
    Above the **TextFragment** element, **Offset** = –50; **Element border** = **Bottom**;

12. Match the FlexiLayout with the test images.
    You will notice that the program correctly detects the table element on all the images, but on some of the images the **Quantity** column includes redundant text from the preceding column.
    To separate the good data in the **Quantity** column from irrelevant data, we will specify the type of context for this column.
13. Open the **Properties** dialog box of the table element **InvoiceTable**.
    Click the **Columns** tab, open the **Properties...** dialog box of the **Quantity** column, and select the **Check content type** option. In **Type of column content** select **Number**.



Properties of column Quantity

○ Detect by keyword in name:

Qty

□ Whole words only

□ Multi-line name

Missing keyword penalty: 0.97

● Use found element as column name:

SearchElements.TableHeader.kwQuantity

☑ Check content type

Type of column content: Number

Regular expression:

Missing content penalty: 0.98

□ Absolute area constraints:

Left boundary: 0

Right boundary: INF

14. Now if you match the FlexiLayout on the images, you will see that the **Quantity** contains only digits, which is what we need.
15. Similarly, specify the **Currency** context for the **UnitPrice** and **Total** columns.
16. To specify the location of the **InvoiceTable** block, select the **Source element** option. Click "**...**" and specify **InvoiceTable** as the source element.

Step 9. TotalAmount field, SumGroup group element

Once you have analyzed the test documents, you will notice that the **Total Amount** field occurs only together with the title on the last page below the table. Additionally, the field itself is either on the same level as the title or below the title. Since the field contains only one line, we will create a **Character String** element to search for it.

To specify the settings common to the field and its title, we will use a group element.

To create a **SumGroup** element:

1. Create an element of the **Group** type and name it **SumGroup**.
2. Click the **Advanced** tab and in the **Advanced pre–search relations** field, specify additional search constraints for all the elements in the group: *Search for the object below the last detected table*. In the FlexiLayout language, this constraint can be written as follows:

```
If not InvoiceTable.LastSubTable.Body.IsNull then
Below: InvoiceTable.LastSubTable.Body.Rect.Bottom;
```

To search for the keywords in the **TotalAmount** field, use a **SumGroup.kwTotal** element of the **Static Text** type. See **Step 9.1** for detailed instructions.

To search for the one–line **TotalAmount** field, use a **SumGroup. TotalAmount** element of the **Character String** type. See **Step 9.2** for detailed instructions.

Step 9.1. Total Amount title, kwTotal element

*This title occurs on the images as "**Total Chf,**" "**Carried Over,**" or "**Total.**" It is located below the table (this constraint has already been specifies for the **SumGroup** element) and always occurs on the last page.*

To create a **kwTotal** element:

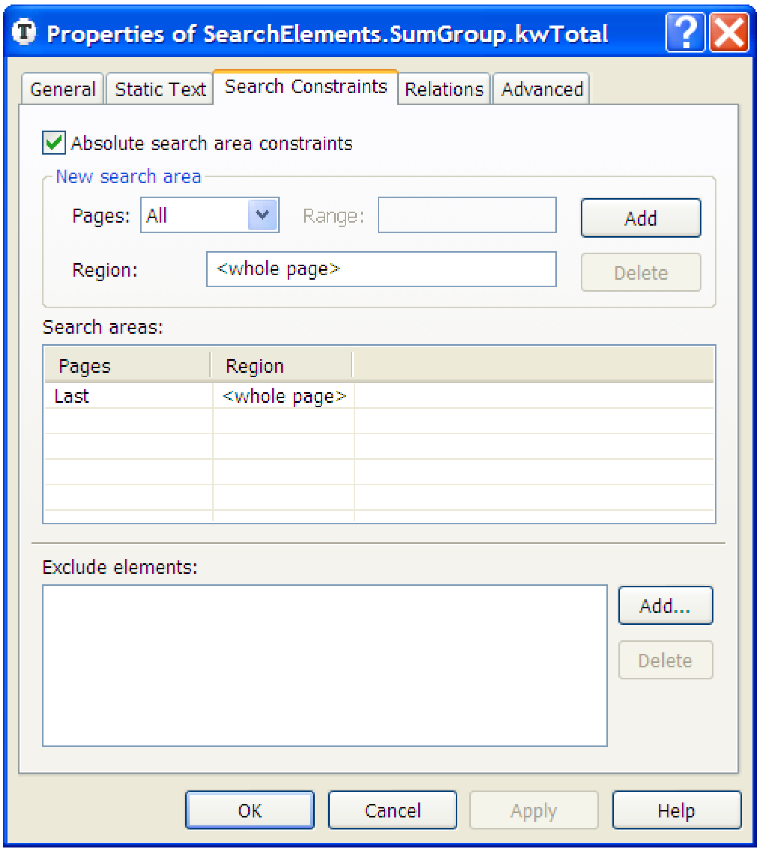
1. In the **SumGroup** element, create an element of the **Static Text** type and name it **kwTotal**.
2. Click the **Static Text**. tab.
3. In the **Search Text** field, specify the text to search: **TotalChf|CarriedOver|Total**.
4. Set **Max error percentage** to 20.
5. Click the **Search Constraints** tab and limit the search area to the last page:

   Select **Absolute search area constraints**

   In the **Pages** field, select Last from the drop–down list

   Do not select any other options

   Click the **Add** button. The search area will appear in the **Search areas** list



☑ Absolute search area constraints

Pages: All

Region: <whole page>

Search areas:

Pages	Region
Last	<whole page>

6. Click the **Advanced** tab.

7. Among the many detected elements that contain the specified text, we must select the one that is located closest to the bottom border of the table. This location allows us to influence the quality of the hypotheses. In **Advanced post-search relations** specify additional parameters for the hypothesis: *If a non-null hypothesis has been generated, calculate the distance between the upper border of the detected element and the bottom border of the last table detected in the document. Then check if this distance belongs to the specified fuzzy interval.* In the FlexiLayout language, this condition can be written as follows:

```
If not IsNull then
{ FuzzyQuality: Top – InvoiceTable.LastSubTable.Body.Rect.Bottom, {0,0,0, 50000}*dt;
}
//This check means that the greater the distance between the borders, the greater the penalty coefficient.
```

Step 9.2. Total Amount field, TotalAmount element

We will search for the **TotalAmount** field both to the right of and below the title.

To create a **TotalAmount** field:

1. In the **SumGroup** element, create an element of the **Character String** type and name it **TotalAmount**.
2. Click the **Character String** tab.
3. Specify the alphabet:
   ',–.0123456789OSZosz
4. Set **Percentage of non-alphabet characters** to 30%.
5. In the **String length** field, specify the following fuzzy interval {–1, 1, 20, INF}, which assesses the string length. We assume that all the possible values will fall into the interval of 1 to 20 characters. Any hypothesis outside this interval will be penalized
6. Do not impose any additional constraints and keep the default settings.
7. Click the **Search Constraints** tab and exclude the region of the hypothesis for the **kwTotal** element from the search area.
8. Click the **Advanced** tab.
9. In **Advanced pre-search relations** specify additional search conditions: If the **kwTotal** title has been detected, specify the search area as an array of rectangles: one rectangle to the right of the field title and one rectangle below the field title with some offset. Otherwise, do not search for the image object. In the FlexiLayout language, this condition can be written as follows:

```
If not kwTotal.IsNull Then // If the title kwTotal has been detected
{
WholeWordMOde: true;
MaxGapInLine: 30dt;

RectArray DataRegion;

Let r1= Rect (kwTotal.Rect.Right, kwTotal.Rect.Top –20dt, kwTotal.Rect.Right + 1800dt, kwTotal.Rect.Bottom + 50dt);
// Limits the search area to a rectangle to the right of the field title.
Let r2 = Rect (kwTotal.Rect.Left – 100dt, kwTotal.Rect.Bottom, kwTotal.Rect.Right + 100dt, kwTotal.Rect.Bottom + 200dt);
// Limits the search area to a rectangle below the field title.

DataRegion = RectArray (r1);
DataRegion.Add (r2);

RestrictSearchArea (DataRegion);
}
else Dontfind(); // Otherwise, do not search for the object.
```

10. The search area of the field value is limited to two rectangles, one to the right of and the other below the field title. If the field value is located to the right of the field title, there is some text below the title on most of the images. Since on most of the images this redundant text is included into the search area of the **TotalAmount** field, the program will create a hypothesis for this text. To influence the quality of this hypothesis (i.e. to lower its quality), in **Advanced post-search relations** specify an additional check: *The lower the field value below the field title, the greater the penalty.* In the FlexiLayout language, this check can be written as follows:

```
If not IsNull then
{FuzzyQuality: 400dt – width, {–50000,0,0, 50000}*dt;
//Hypotheses for elements with region 400 dots wide will have the maximum quality
if width < 50dt then Quality: 0;
// Elements less than 50 dots are not considered
if Top > kwTotal.Rect.Top then
{ FuzzyQuality: Top – kwTotal.Rect.Top, {0,0,0, 20000}*dt;
}
}
//This check means that the higher the difference, the greater the penalty.]
```

11. Match the FlexiLayout on the images to make sure that the program can reliably detect the element on all the images where it occurs.

12. Specify the location of the **TotalAmount** block as the rectangular region of the **TotalAmount** element expanded by 5 dots vertically and horizontally. To do this, select **Region expression** and specify the following expression:

```
Rect outputRect;
outputRect = SumGroup.TotalAmount.Rect;
IsNull = SumGroup.TotalAmount.IsNull;
OutputRegion = outputRect;
OutputRegion.Inflate (5dt, 5dt);
```

Step 10. Company field, CompanyGroup group element, Company element

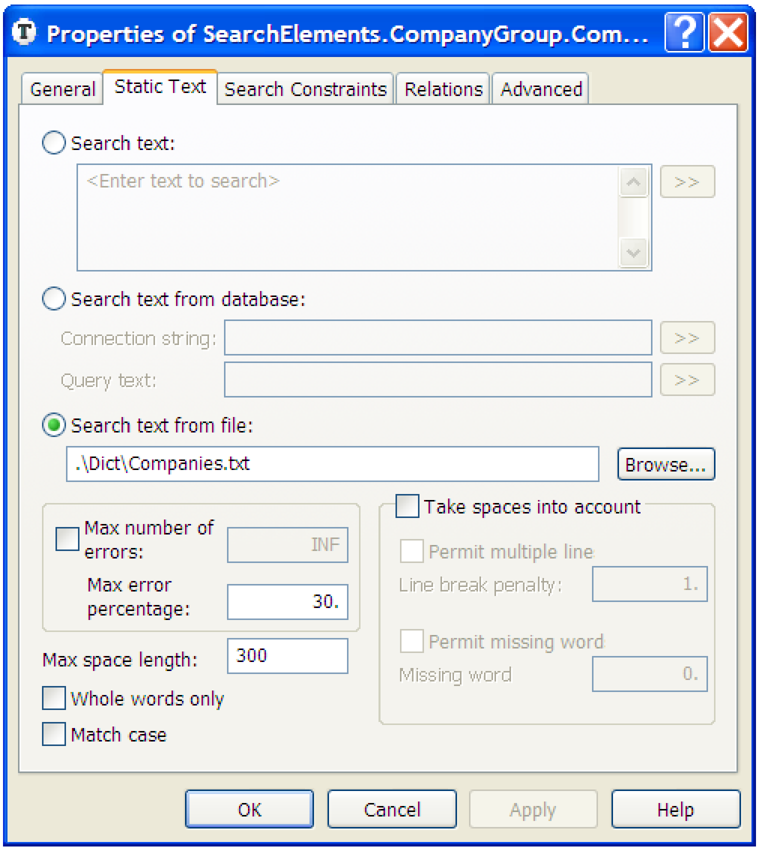
Once you have analyzed the images, you will notice that the **Company** field is always located in the first third of the first page of each document.

To search for the **Company** field, use an element of the **Static Text** type which is part of a **CompanyGroup** element of the **Group** type.

To create a **CompanyGroup** element, create an element of the **Group** type and name it **CompanyGroup**. Keep the default settings.

To create a **Company** element:

1. In the **CompanyGroup** element, create an element of the **Static Text** type and name it **Company**.
2. Click the **Static Text** tab.
3. Select **Search Text from File** and specify the path to the **Companies.txt** file.
   Note: The **Companies.txt** files is located in
   <disk name>:\\\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\9.0\Samples \FlexiLayout Studio\Invoice\Dict.



○ Search text: <Enter text to search>

○ Search text from database:

Connection string:

Query text:

● Search text from file: .\Dict\Companies.txt

□ Max number of errors: INF

Max error percentage: 30.

Max space length: 300

□ Whole words only

□ Match case

□ Take spaces into account

□ Permit multiple line

Line break penalty: 1.

□ Permit missing word

Missing word 0.

4. Click the **Advanced** tab.
5. In **Advanced pre-search relations**, specify an additional search constraint: *Search for the image object in the first third of the first page of the document.* In the FlexiLayout language, this constraint can be written as follows:

   Above: Page(1).RectGlobal.Top + Page(1).RectGlobal.Height/3;

6. Match the FlexiLayout to make sure that it reliably detects the element on all the images where it occurs.
7. When you review the matching results, you will notice that the program failed to find the name of the company on some of the documents. This is becuase on some of the documents the text is written on two lines, whereas we specified one-line text in the properties of the element. To correct this, make the following changes

   Open the **Companies.txt** file and specify the comapny name as **YOUR SERVICE** with spaces. This means that the program should look for a phrase which need not be written on one line.

   Open the **Properties** dialog box of the **Company** element and click the **Static Text** tab.

Select **Permit multiple lines** to specify that the phrase can be written on more than one line.

Set **Line break penalty** to 1. This will retain the quality of the generated hypothesis even if the text is written on more than one lines. A smaller value would mean that the hypothesis should be penalized.

8. Match the FlexiLayout to maker sure that the program reliably detects the element on the test images.
9. Specify the location of the **Company** block as the rectangular region of the **Company** element expanded by 5 dots vertically and horizontally. To do this, select **Region expression** and specify the following expression:

```
Rect outputRect;
outputRect = CompanyGroup.Company.Rect;
IsNull = CompanyGroup.Company.IsNull;
OutputRegion = outputRect;
OutputRegion.Inflate (5dt, 5dt);
```

Step 11. Exporting the FlexiLayout and adding it to a FlexiCapture template

You can export your FlexiLayout into an *.afl file and add it to an ABBYY FlexiCapture template.

Note: For detailed instructions, see **Sample 1, Step 21 and Step 22**.