




FlexiLayout Studio 9.0

Tutorial Sample 1

Contents

Introduction	3
Step 1. Creating a new project.....	3
Step 2. Adding images to the batch	4
Step 3. Setting the FlexiLayout properties	6
Step 4. Pre-recognition	7
Step 5. Viewing images and pre-recognition results	8
Step 6. Analyzing pre-recognition results and selecting reference elements	10
Step 7. Creating a form identifier	10
Step 8. Testing the identifier element.....	12
Step 9. Adjusting the properties of the identifier element.....	12
Step 10. Describing the YOUR PLANET NAME field	13
Step 11. Describing the YOUR PLANET NAME field: PlanetNameHeader element.....	14
Step 12. Describing the YOUR PLANET NAME field: PlanetName element.....	15
Step 13. Testing the YOUR PLANET NAME field.....	19
Step 14. Describing the YOUR PLANET NAME field: PlanetName block.....	20
Step 15. Describing the NAME field.....	21
Step 16. Describing the YOUR SPACESHIP NUMBER field.....	22
Step 17. Describing the DATE YOU ARRIVED AT THE EARTH field	23
Step 18. Describing the YOUR IDENTITY NUMBER ON THE PARTY field	25
Step 19. Describing the ANY TEXT field	25
Step 20. Describing the YOUR PHOTO IN FANCY DRESS field	27
Step 21. Exporting the FlexiLayout.....	30
Step 22. Opening the FlexiLayout in ABBYY FlexiCapture.....	30

Introduction

 **Important!** For the sake of simplicity, a one–page document is used in this sample.

This Tutorial describes the creation of a FlexiLayout for a simple registration form. You will learn how to create blocks and elements, set simple search constraints and describe field locations. The FlexiLayout Studio project with the test images and a ready–made FlexiLayout can be found in

<disk name>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\9.0\Samples\FlexiLayoutStudio\Halloween Form.

This sample document is a typical example of a registration form for which a FlexiLayout must be created. You will go through all the steps required to create a FlexiLayout that will detect the following elements: **Static Text, Character String, Barcode, Text Fragment, Date, Separator, White Gap** and **Object Collection**.

The ultimate goal is to create a FlexiLayout which will enable ABBYY FlexiCapture to locate the following fields on all the test pages:


NAME
YOUR PLANET NAME
YOUR SPACESHIP NUMBER
DATE YOU ARRIVED AT THE EARTH
YOUR IDENTITY NUMBER ON THE PARTY
YOUR PHOTO IN FANCY DRESS

Additionally, the FlexiLayout must enable FlexiCapture to find a text field in which the guest has entered the information he or she managed to find about Halloween. Unlike all the above fields, the exact title of this field is not known in advance.

HALLOWEEN REGISTRATION FORM

NOTE: Please remember that this form should be completed by extraterrestrials only

YOUR PHOTO IN FANCY DRESS




NAME: BIG-TROUBLE-AND-EVEN-MORE-THAN-YOU-THINK

YOUR PLANET NAME: PLUTO

YOUR SPACESHIP NUMBER: 000 00 WD 01

DATE YOU ARRIVED AT THE EARTH: 10/ 10/ 2005

YOUR IDENTITY NUMBER ON THE PARTY IS:



About some customs:

The Jack-o-lantern custom probably comes from Irish folklore. As the tale is told, a man named Jack, who was notorious as a drunkard and trickster, tricked Satan into climbing a tree. Jack then carved an image of a cross in the tree's trunk, trapping the devil up the tree. Jack made a deal with the devil that, if he would never tempt him again, he would promise to let him down the tree.

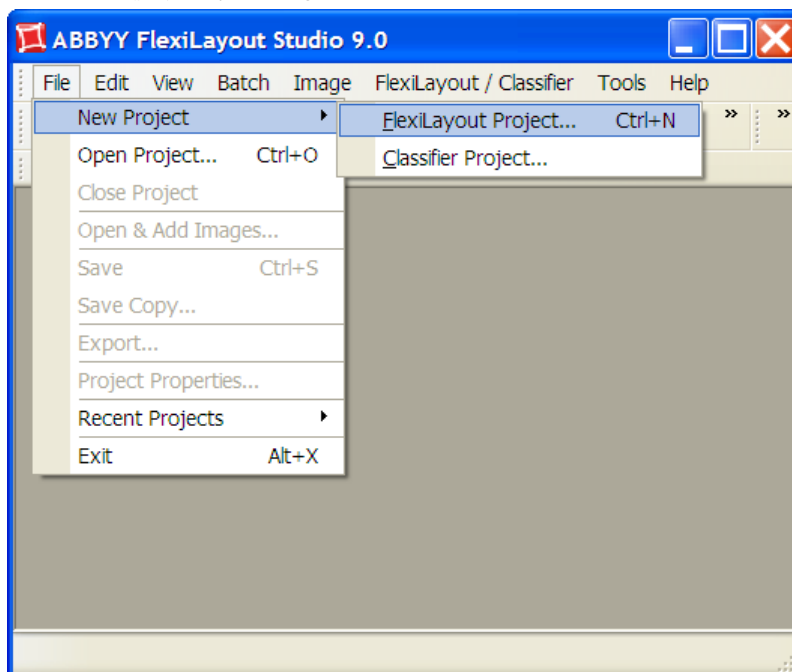
According to the folk tale, after Jack died, he was denied entrance to Heaven because of his evil ways, but he was also denied access to Hell because he had tricked the devil. Instead, the devil gave him a single ember to light his way through the frigid darkness. The ember was placed inside a hollowed-out turnip to keep it glowing longer.

The Irish used turnips as their "Jack's lanterns" originally. But when the immigrants came to America, they found that pumpkins were far more plentiful than turnips. So the Jack-O-Lantern in America was a hollowed-out pumpkin, lit with an ember.

Step 1. Creating a new project

 **Note:** A project is a set of files created by ABBYY FlexiLayout™ Studio which contains all the data pertaining to a FlexiLayout.

1. Create a new folder and name it **Sample1**.
2. Run ABBYY FlexiLayout Studio.
3. Create a new project by selecting **New Project** in the **File** menu.




4. In the dialog box that opens, type the name for the project: **Sample1**.

After you create the project, the **Halloween Form** folder will contain:

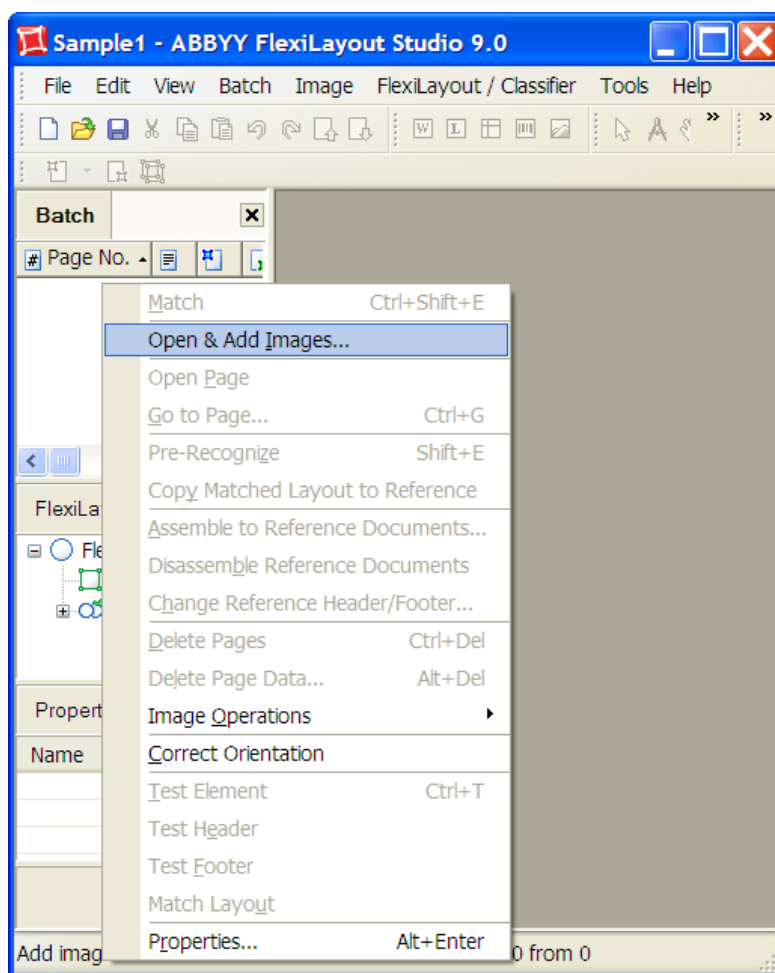
- **Sample1.fsp** – the file of your project,
- **Sample1Batch** – the default folder where all images added to the batch are stored,
- **Sample1Templates** – the default folder where the FlexiLayout is stored.

Step 2. Adding images to the batch

 **Note:** A batch is a set of test images (pages) to be used for testing and adjusting a FlexiLayout.

Once the new project has been created, you must add the test images to the batch. The images will be used to test and adjust the FlexiLayout.

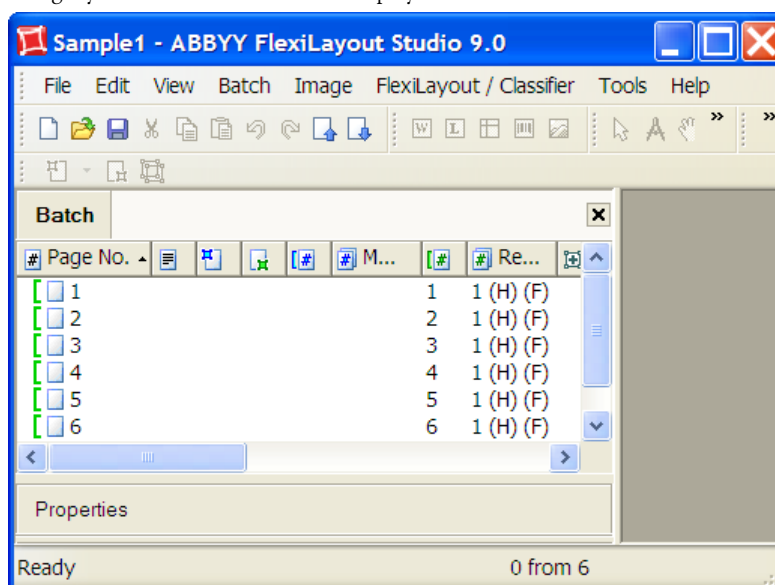
1. Click on the **Batch** tab in the main program window.
2. Select the **Open & Add Images** command in the **File** menu or in the local menu.




3. In the **Open Images** dialog box, select the **Document per file** option and specify the test image files.

Sample 1 images can be found in
<disk name>:\Documents and Settings\All Users\Application Data\ABBYY\FlexiCapture\9.0\Samples\FlexiLayoutStudio\Halloween Form\Halloween FormBatch.

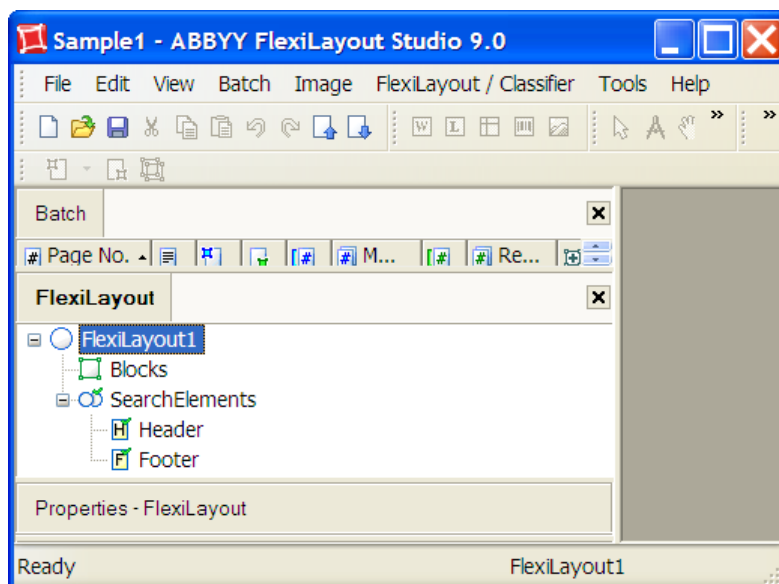
The test images you add to the batch will be displayed in the **Batch** window.




 **Note:** When adding images, you can specify how reference assembly must be performed. By selecting Document per file, you tell the program that each image is a separate document. For more information about reference assembly, see the **Reference document assembly** section of the help file.

Step 3. Setting the FlexiLayout properties

By default, the program has named the new FlexiLayout **FlexiLayout1** (see the **FlexiLayout** tab in the **FlexiLayout Properties** dialog box).



 **Note:** The error icon next to the **SearchElements**, means that no elements have so far been created in the FlexiLayout.

We recommend renaming the FlexiLayout and giving it a meaningful name for the sake of convenience.

To set the FlexiLayout properties (these also include the name of the FlexiLayout):

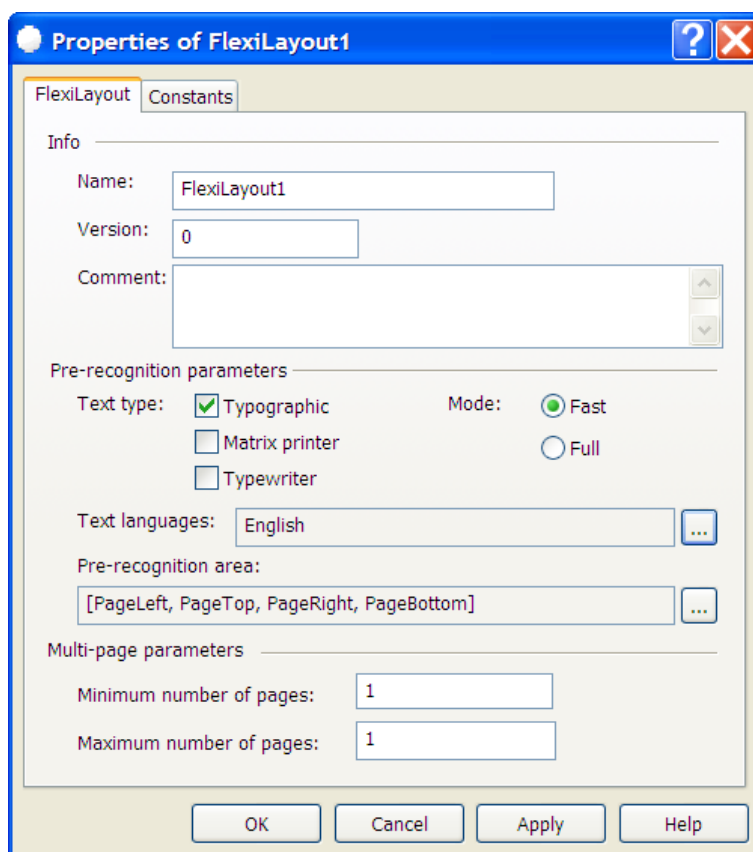
1. Double-click on the name of the FlexiLayout (i.e. **FlexiLayout1**) or right-click on the FlexiLayout and select **Properties** in the local menu.
2. In the Name field, type a new name for the FlexiLayout, e.g. RegistrationForm.
3. Specify the pre-recognition parameters:

From the Text language list, select English as the pre-recognition language (the documents contain text in English).


Select **Typographic** in the **Text type** group (this is the default setting), assuming that in this particular case neither a dot-matrix printer nor a typewriter will be used to fill in the form.

Mode. You can select either **Fast** or **Full** pre-recognition mode, depending on the quality of the images to be processed. Since the sample images are of good quality, contain no noise and almost no small print, select the default **Fast** mode. This will speed up pre-recognition, both when creating the FlexiLayout in FlexiLayout Studio and when processing it in FlexiCapture.

Multi-page parameters – Our document contains one page. Therefore, specify 1 for **Minimal number of pages** and 1 for **Maximal number of pages**.



There is no need to specify the beginning and end of documents, so you can delete the predefined **Header** and **Footer** elements from the list of elements of your FlexiLayout.

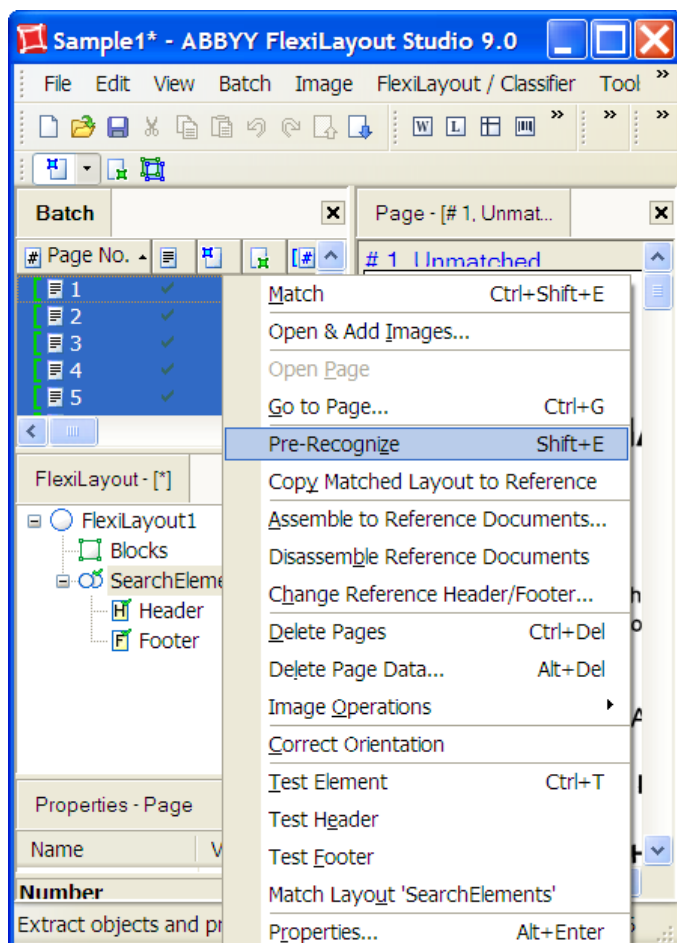
 **Note:** You can add **Header** and **Footer** elements to a FlexiLayout to help the program detect the beginning and end of documents. If the program fails to find both the **Header** and **Footer** elements when matching the FlexiLayout, it will use the maximum number of pages specified in the FlexiLayout for document separation. For the document in this example, we specified 1 as the maximum number of pages.

Step 4. Pre-recognition

Before you start creating form elements, you need to know which objects on the form can be used as "signposts" when looking for fields (blocks). These are usually pictures and/or text fragments that are consistently detected during pre-recognition.

To start pre-recognition:

1. Select all the images in the batch.
2. Select the **Prerecognize** command in the **Batch** menu or in the local menu of each image.



Note: If you change images in the batch or add new images, you will need to restart the pre-recognition procedure. If an image has not been pre-recognized, it will be pre-recognized when matching it with its FlexiLayout.

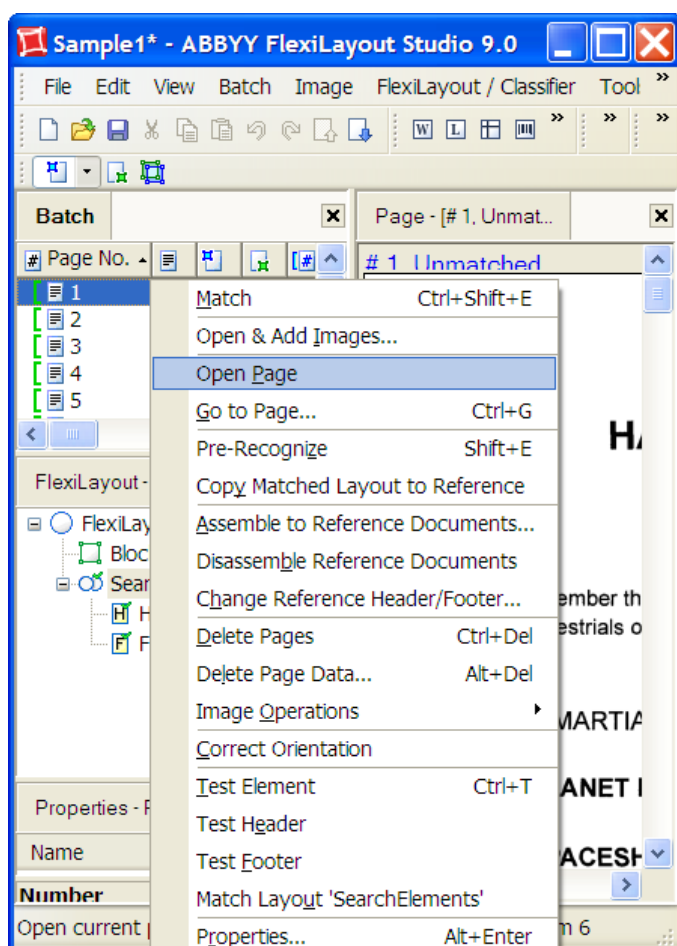
Step 5. Viewing images and pre-recognition results

Open the first page in the batch:

Double-click on the corresponding line in the **Batch** window.

or

Select the **Open Page** command in the **Batch** menu or in the local menu of the batch.

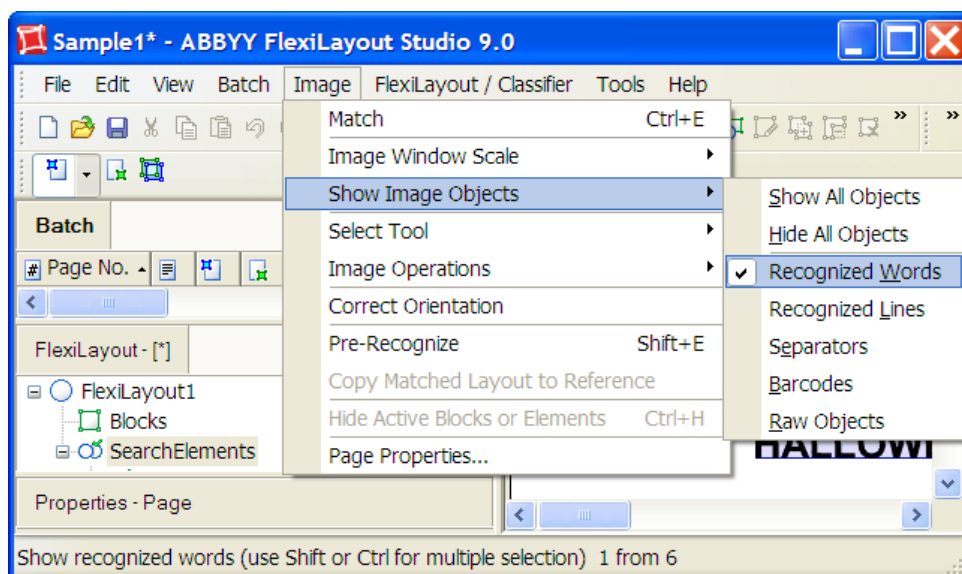


The image will be displayed in the **Image** window in **Matched Layout** mode.

By default, the pre-recognition results are not displayed on the image.

To view the pre-recognition results:

Select the **Show Image Objects** command in the **Page** menu and then select the type of image object you wish to see,



or

Click the corresponding icon on the **Extracted Objects** toolbar.



 **Note:** You can also view all the objects by selecting the Show All Objects command.

Step 6. Analyzing pre-recognition results and selecting reference elements

When analyzing the pre-recognition results, you might wish to view different types of objects and images separately.

First, find the objects which have been consistently detected during pre-recognition on each of the test images and which the program may use as reference elements when looking for fields.

Once you have viewed the pre-recognition results for all the test images, you will notice that:

1. The program has consistently detected all the Text objects corresponding to the titles of the fields, and all the Barcode objects and Separator objects around the photograph. These objects can be used as reference elements when looking for the blocks.
2. On images 3, 5 and 6 the program has detected separators that are missing on the other images. These objects cannot be used as reference elements when looking for the blocks.
3. On images 1 and 6 the program has failed to recognize all parts of the photograph as Image objects of the Picture type. This means that looking just for a Picture object the program will not find the entire photograph, and you will need to use additional methods. (They will be described in Step 20: Describing the Photograph field.)

Now you can create and set up the reference elements and specify the search constraints for blocks corresponding for fields to be search.

Step 7. Creating a form identifier

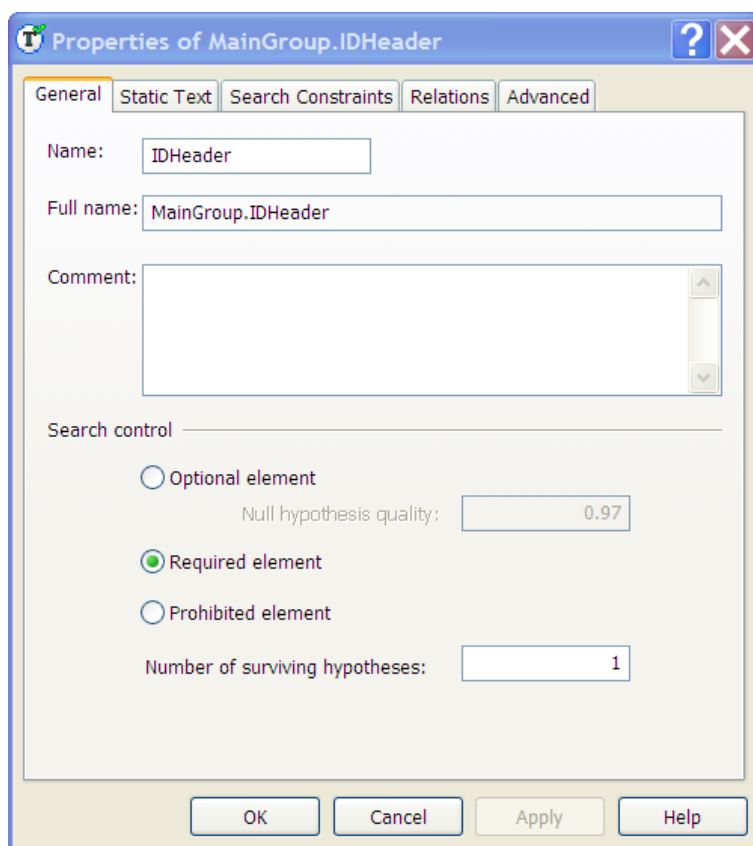
When processing semi-structured forms in ABBYY FlexiCapture, one would normally wish to exclude forms not belonging to the current type. One way to identify a form is to mark at least one element as required. A required element must be consistently detected on all the forms of a given type (otherwise, the program will not be able to match the form and its FlexiLayout). In this particular case, the form heading (HALLOWEEN REGISTRATION FORM) will make a good identifier element, since it contains distinct text that can be easily read by the OCR engine.

 **Note:** You can specify an identifier element or set of elements in a predefined compound **Header** element (not described in this tutorial).

The form heading will be used solely to identify the form as belonging to the given type. In the FlexiLayout, specify the form heading as an element of the **Static Text** type.

To create an ID element:

1. Click on the **FlexiLayout** tab in the program main window.
2. Select **SearchElements** in the FlexiLayout tree.
3. Select the Elements command in **FlexiLayout>Add element>Static Text** or in the local menu of the element.
4. In the **Name** field, type a name for the element, e.g. **IDHeader**.
5. Clear the **Optional element** box on the **General** tab to make the form heading a required element.



Properties of MainGroup.IDHeader

General Static Text Search Constraints Relations Advanced

Name: IDHeader

Full name: MainGroup.IDHeader

Comment:

Search control

☐ Optional element
Null hypothesis quality: 0.97

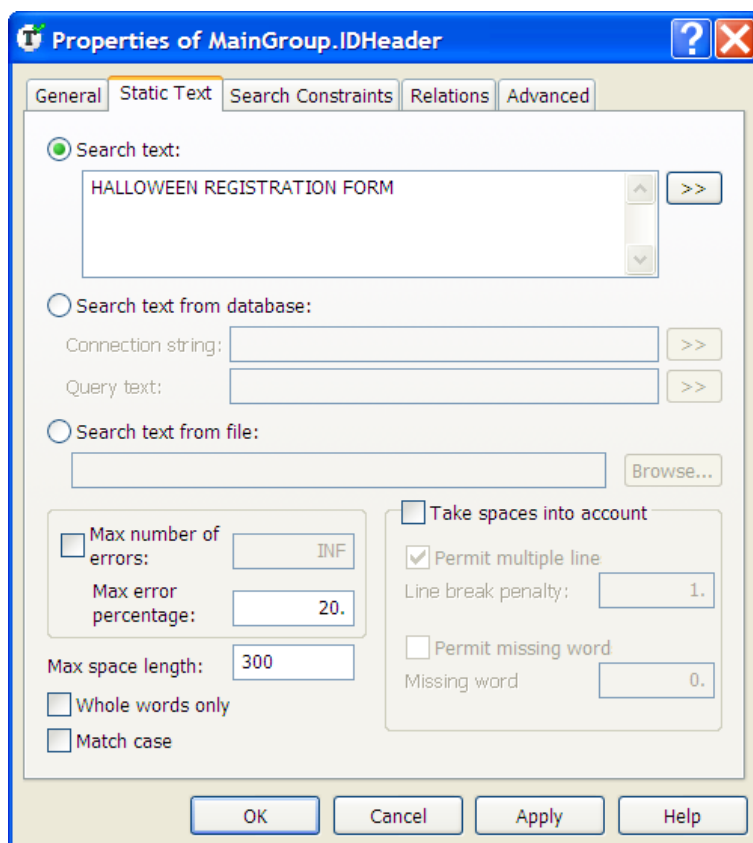
☒ Required element

☐ Prohibited element

Number of surviving hypotheses: 1

OK Cancel Apply Help

6. Click on the **Static Text** tab.



Properties of MainGroup.IDHeader

General Static Text Search Constraints Relations Advanced

☒ Search text:

HALLOWEEN REGISTRATION FORM

☐ Search text from database:

Connection string:

Query text:

☐ Search text from file:

Browse...

☐ 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: 1.

☐ Permit missing word

Missing word: 0.

OK Cancel Apply Help

7. In the **Search text** field, type the text to find: **HALLOWEEN REGISTRATION FORM**. Judging by the first image in the batch, one can assume that the form heading is written in one line. Therefore, you can type the heading without spaces to speed up looking for single-line static text.


- Set the maximum number of errors that may occur in the found text (either in percentage points or as a number). In this particular case we recommend setting the **Max error percentage** at 20, allowing 5 errors for 25 characters of the form heading.

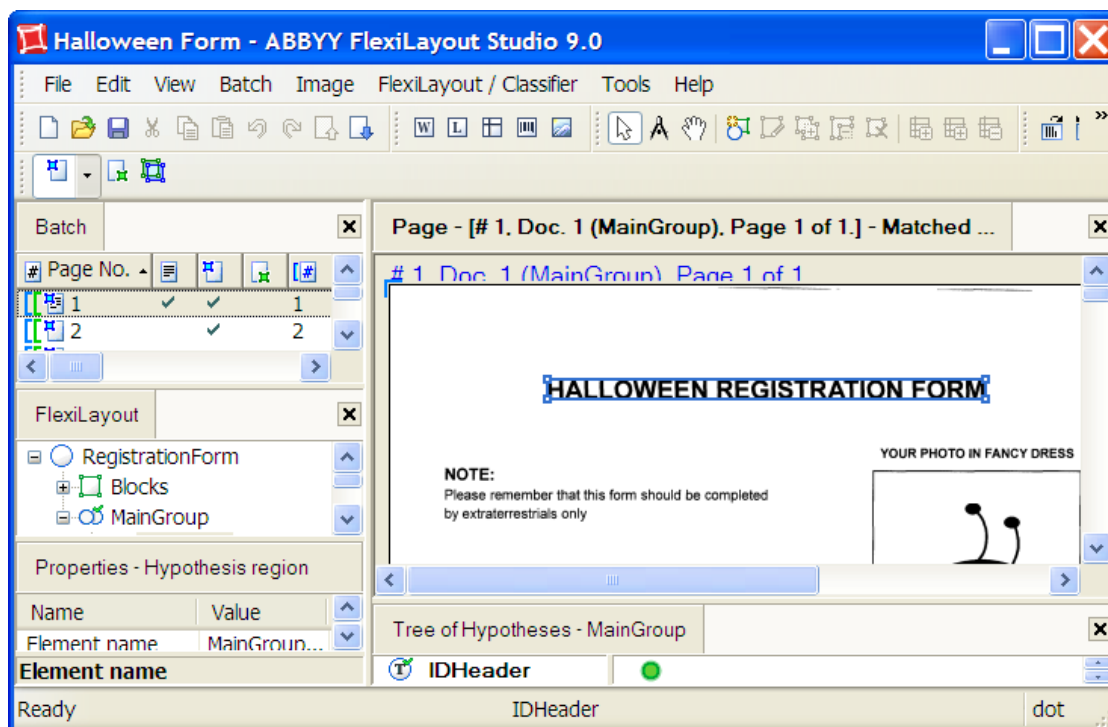
 **Note:** The optimal percentage of allowed errors can only be found by trial and error.

Step 8. Testing the identifier element

To check that the program can reliably detect the image object corresponding to identifier element **IDHeader**, try matching the FlexiLayout with each image in the batch:

- Open each image one by one.
- Select the **Match FlexiLayout & Show Hypotheses** command in the **Page** menu or in the local menu of each image.

If the image and its FlexiLayout have been matched successfully, the program draws a blue frame around the detected element and the hypothesis which the program formulates about the **IDHeader** element will be marked with  in the **Tree of Hypotheses**. If you click on the hypothesis, you will see the properties of the hypothesis in the **Properties** window.

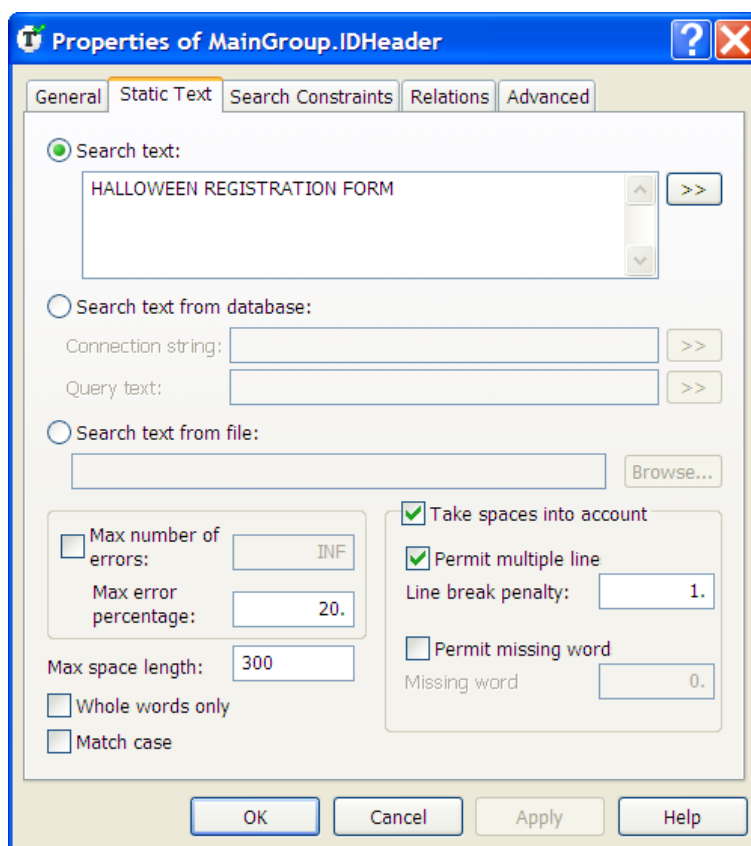


Once you have examined the FlexiLayout matching results for all the images, you will see that the program has failed to formulate a hypothesis for the **IDHeader** element on image 2. This is because the form heading is written in two lines on this image, whereas you described the **IDHeader** as single-line text.

Step 9. Adjusting the properties of the identifier element

For the identifier element to be detected consistently on all the forms, you need to adjust its properties:

- Open the element's **Properties** dialog box.
- Click on the **Static Text** tab.



3. In the **Search text** field, type the text to find, preserving the spaces: **HALLOWEEN REGISTRATION FORM**. This means that the program will look for a phrase which may not be written in one line.
4. Select **Allow for spaces** and **Permit multiple lines** to specify that the search phrase may be written on more than one lines.
5. Set the **Line break penalty** to 1. This will preserve the quality of the hypothesis when it goes to the next line. If you set a smaller value, the hypothesis will be penalized.

Now if you try matching the FlexiLayout and the images, the program will detect the identifier element on all the test images.

Step 10. Describing the YOUR PLANET NAME field

The field titles are detected reliably on all the test images and do not repeat on each given form. The order of the fields is the same on all the forms, even though their exact locations may vary from form to form. This means that for all the fields you can:

1. Create **Static Text** elements to look for the title of the field.
2. Use the field title to locate **Character String**, **Text Fragment**, **Barcode** etc. elements, which will then be used to look for the corresponding fields.
3. Create blocks (which correspond to fields) and specify their location as coinciding with that of the corresponding reference elements.

To describe the location of the block corresponding to the **YOUR PLANET NAME** field:

1. Create an element of the **Static Text** type and name it **PlanetNameHeader**. This element will correspond to the title of the field **YOUR PLANET NAME**.
2. Create an element of the **Character String** type and name it **PlanetName**. This element will correspond to the field **YOUR PLANET NAME** proper.
3. Create a block of the **Text** type and name it **PlanetName**. This block will correspond to the field **YOUR PLANET NAME**.

Step 11. Describing the YOUR PLANET NAME field: PlanetNameHeader element

Even though the **YOUR PLANET NAME** field is preceded by the **NAME** field, the FlexiLayout must first search for the field **YOUR PLANET NAME**. This will exclude the ambiguity which may arise if the FlexiLayout looks for the **NAME** field first. As the word **NAME** can be found in both field titles, the program will formulate several hypotheses if you instruct it to look just for the word **NAME** without any additional constraints. To avoid this ambiguity, the program must first look for the **YOUR PLANET NAME** field and only then for the **NAME** field. In Step 15 you will impose additional constraints on the **NAME** field.

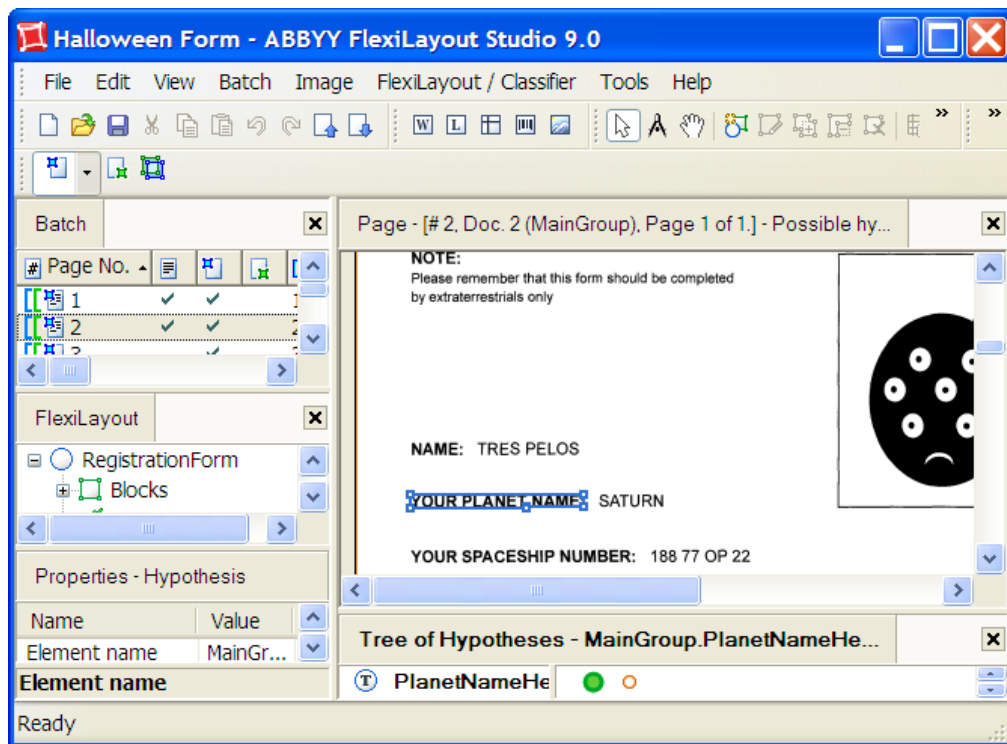
To create a **PlanetNameHeader** element:

1. Create an element of the **Static Text** type and name it **PlanetNameHeader**.
Do not clear the **Optional element** box which is selected by default. Unlike the form heading, which is used as a form identifier and for this reason was specified as a required element, all the other form elements are assumed to be optional. This will allow the program to formulate null hypotheses even for those elements for which it fails to find the corresponding objects on the images.

The screenshot shows a dialog box titled "Properties of MainGroup.PlanetNameHeader". It has five tabs: "General", "Static Text", "Search Constraints", "Relations", and "Advanced". The "General" tab is selected. Inside the "General" tab, there are three text input fields: "Name:" with the value "PlanetNameHeader", "Full name:" with the value "MainGroup.PlanetNameHeader", and "Comment:" which is empty. Below these is a "Search control" section. It contains three radio buttons: "Optional element" (which is selected), "Required element", and "Prohibited element". To the right of the "Optional element" radio button is a text box labeled "Null hypothesis quality:" containing the value "0.97". Below the radio buttons is another text box labeled "Number of surviving hypotheses:" containing the value "5". At the bottom of the dialog are four buttons: "OK", "Cancel", "Apply", and "Help".

2. Click on the **Static Text** tab.
3. In the **Search text** field, type the text to search. Since the field title is written in one line on all the test images, you can type it without spaces: **YOURPLANETNAME**.
4. Set the **Max error percentage** to 20, since the field title has more than one word and is fairly long.

Try matching the FlexiLayout with the test images and make sure that the program successfully finds the **YOUR PLANET NAME** field on all the images.



Step 12. Describing the YOUR PLANET NAME field: PlanetName element

We assume that the **YOUR PLANET NAME** field always contains only one line. Since, unlike the field title, the contents of this field is not fixed, you must define it as a **Character String** element.

To create the **PlanetName** element:

1. Create an element of the **Character String** type and name it **PlanetName**.

The screenshot shows the 'Properties of MainGroup.PlanetName' dialog box with the 'General' tab selected. The 'Name' field contains 'PlanetName' and the 'Full name' field contains 'MainGroup.PlanetName'. The 'Comment' field is empty. Under the 'Search control' section, the 'Optional element' radio button is selected, with a 'Null hypothesis quality' of 0.97. The 'Required element' and 'Prohibited element' radio buttons are unselected. The 'Number of surviving hypotheses' is set to 5. At the bottom are buttons for 'OK', 'Cancel', 'Apply', and 'Help'.

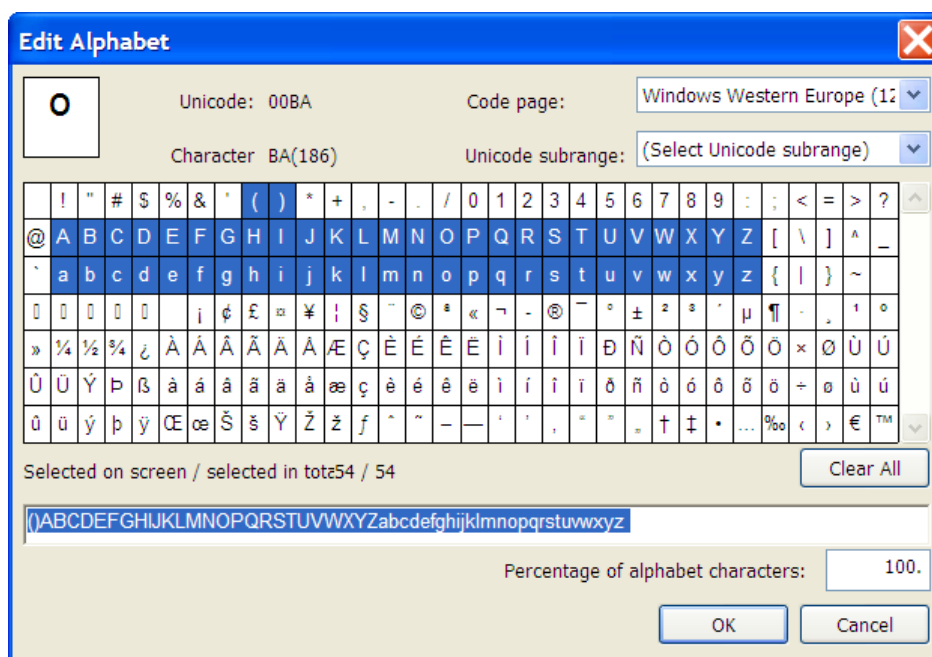
2. Click on the **Character String** tab.

The screenshot shows the 'Properties of MainGroup.PlanetName' dialog box with the 'Character String' tab selected. The 'Regular expression' radio button is unselected, and the 'Alphabets' radio button is selected. Below 'Alphabets' is a table with two columns: 'Character Set' and 'Portio...'. The first row shows '()ABCDEFGHIJKLMNOPQRSTUVWXYZabcd...' and '100.'. To the right of the table are buttons for 'Add...', 'Edit...', 'Delete', and arrows for moving items. Below the table, the 'Allow embedded hypotheses' checkbox is unselected, and the 'Percentage of non-alphabet characters' is set to 10. Further down, the 'Whole words only' and 'Detect words by interword space' checkboxes are unselected. The 'Max space length' is set to 300, and the 'Min interword space' is set to 20. The 'Character count' field shows '{ 0, 3, 35, 100 }' and the 'Word count' field shows '{ -1, -1, INF, INF }'. At the bottom are buttons for 'OK', 'Cancel', 'Apply', and 'Help'.

3. Set the alphabet, i.e. all the characters that may be encountered in the names of planets and their satellites. Judging by the test images, the alphabet consists of English letters and parentheses. To set the alphabet, click the **Add** button and select the required characters in the **Add New Alphabet** dialog box that opens.

Note: The selected characters will be displayed in the **Character Set** column, (**Character String** tab, **Alphabets** field).

- Set the **Percentage of alphabet characters** to 100. This means that the program will take into account only the characters of the current element when looking for the object corresponding to **PlanetName** element.



Note: Generally, you can specify several alphabets for one element. In the **Percentage of alphabet characters** field, specify the maximum percentage of the characters of each alphabet in the element.

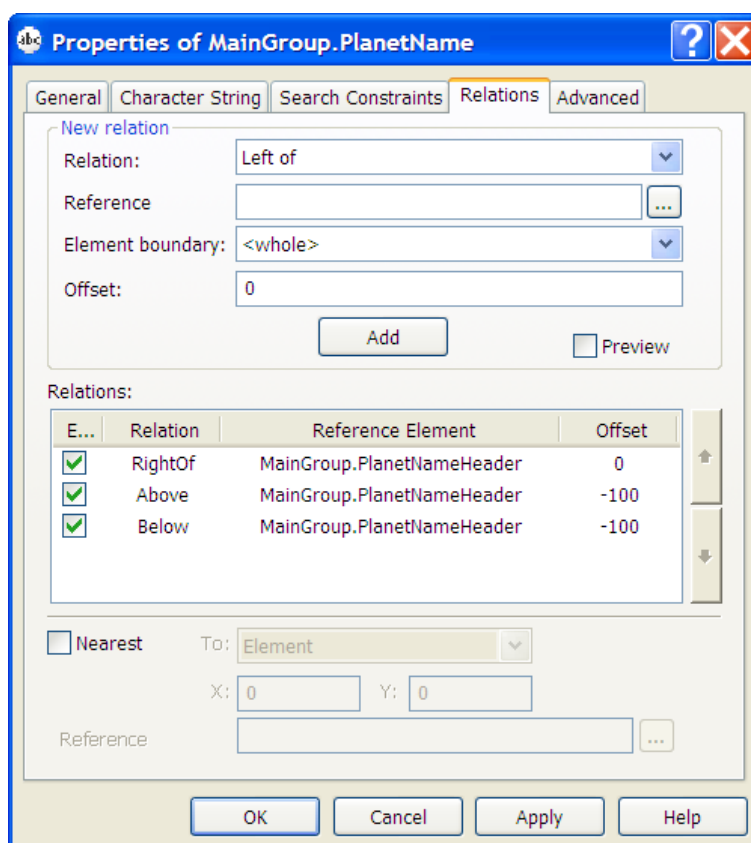
- Clear the **Allow embedded hypotheses** box. This will allow the program to formulate hypotheses which have the maximum length and meet all the search criteria. Otherwise the program may formulate several embedded hypotheses, each consisting of portions of one of the selected alphabets (taking into account the allowed percentage of non-alphabet characters).
- In the **Percentage of non-alphabet characters** field set the maximum percentage of non-alphabet characters to 10%.

Note: This value can only be selected by trial and error and can be changed when adjusting the FlexiLayout.

- In the **String length** field, specify this fuzzy range: {0, 3, 35, 100}. This is an estimation of the length of the string of characters. We assume that the possible names of planets may have from 3 to 35 characters. Any hypotheses outside this range will be penalized.

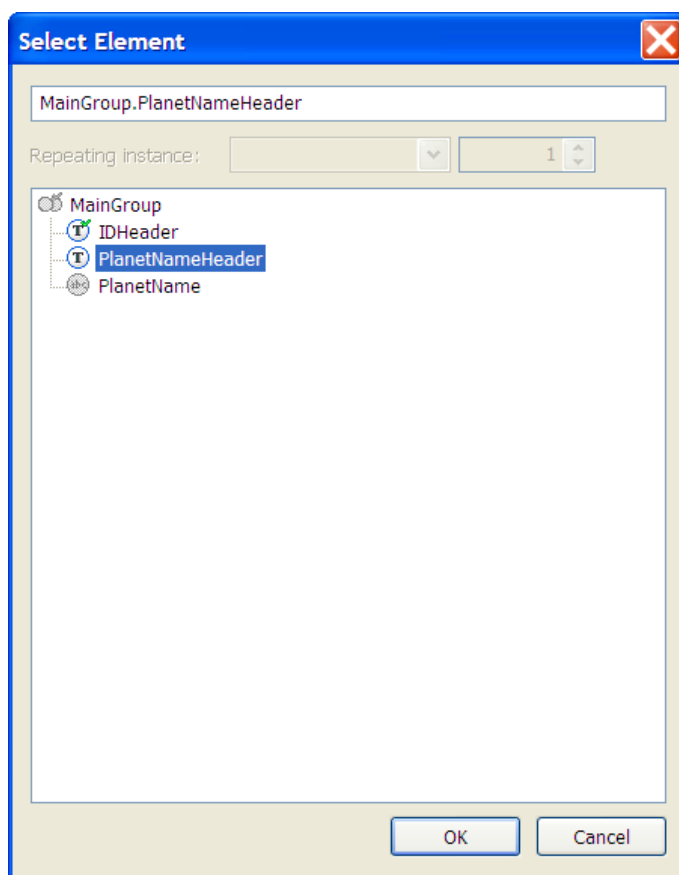
Note: This value can only be selected by method of trial and error and can be changed when adjusting the FlexiLayout.

- In the **Total length of spaces** field, specify a fuzzy range that estimates the total length of all the spaces between the characters in the hypothesis (measured in dots, 1 dot = 1/300 inch, which is the default setting). Do not change the default range of {-1, -1, INF, INF}. This means that there are no constraints on the number of spaces in the hypothesis.
- Click on the **Relations** tab.




Use the **PlanetNameHeader** element as the reference element for the **PlanetName** element. The name of the planet will always be to the right of the field title and will be located on the same level with the title. Therefore, you must specify the location of the element relative to the title and to the title's top and bottom borders. Otherwise, the program will formulate valid hypotheses for any text fragment that consists of characters of the set alphabet, even if it is located higher or lower than the title. To sum up: the program must look for the name of the planet described by the **PlanetName** element to the right of the field title specified as the **PlanetNameHeader** element and located on the same level with the **PlanetNameHeader** element.

10. Specify that the **PlanetName** element is located to the right of the **PlanetNameHeader** element. Select **PlanetNameHeader** in the **Select reference element** window. From the **Relation** drop-down list, select **Rightof** and leave the default value of the **Offset** field equal to 0. Click the **Add** button.




11. Specify that the **PlanetName** element is located not lower than the **PlanetNameHeader** element. Select **PlanetNameHeader** in the **Select reference element** window. In the **Relation** drop-down list, select **Above** and set **Offset** to -100 (this value can only be selected by trial and error). This will give the program some leeway when detecting the position of the element relative to the title. The negative Offset value means that the element is located above the bottom border of the title. Click the **Add** button.
12. Specify that the **PlanetName** element is located not higher than the **PlanetNameHeader** element. Select **PlanetNameHeader** in the **Select reference element** window. In the **Relation** drop-down list, select **Below** and set **Offset** to -100 (this value can only be selected by trial and error). This will give the program some leeway when detecting the position of the element relative to the title. The negative Offset value means that the element is located below the top border of the title. Click the **Add** button.

 **Note:** If you select <whole> in the **Element border** field, positive **Offset** values allow you to specify only the following locations: to the right of the right-hand border of the element, to the left of the left-hand border of the element, higher than the top border of the element, or lower than the bottom border of the element. If you select <top> and <bottom> in the **Element border** field, positive **Offset** values will limit the search area by the top and bottom borders of the element and exclude hypotheses which go beyond the top and bottom borders of the title of the field.

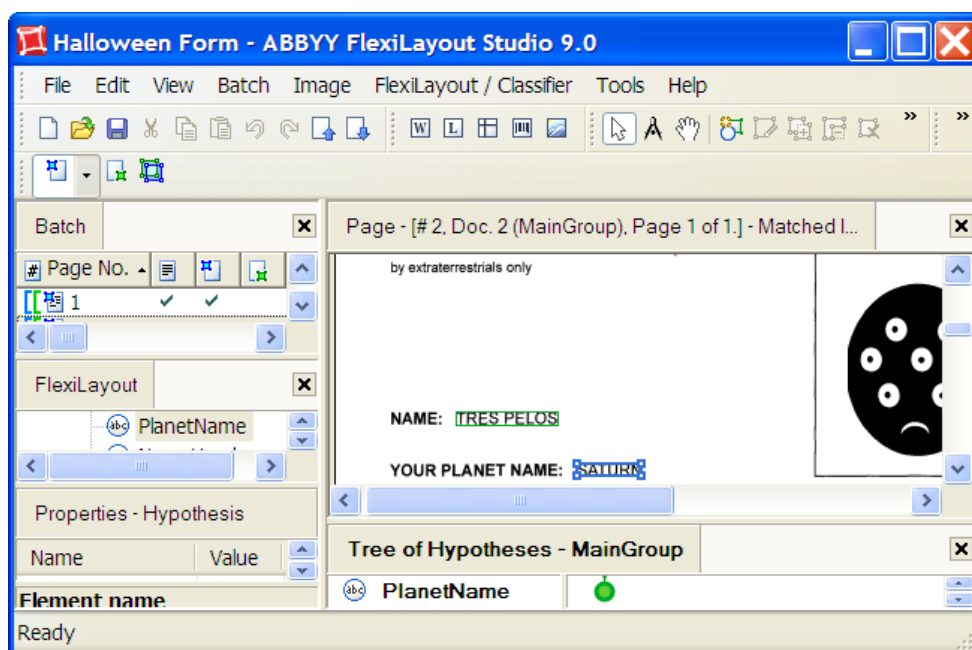
Step 13. Testing the YOUR PLANET NAME field

To check that the program can correctly find the field described by the **PlanetName** element, activate the **Batch** tab in the FlexiLayout Studio main window and try matching the FlexiLayout with each image in the batch:

1. Open an image.
2. Select the **Match FlexiLayout & Show Hypotheses** command in the **Page** menu or in the local menu of each image.

 **Note:** To carry out a fast check, select all the required images in the batch and then select the **Analyze Selected** command in the local menu of the batch. The results can be viewed in the **Image** window in **Matched Layout** mode. To carry out a more thorough check, which takes more time but allows you to view the formulated hypotheses, select the **Match FlexiLayout & Show Hypotheses** command for each open image separately.

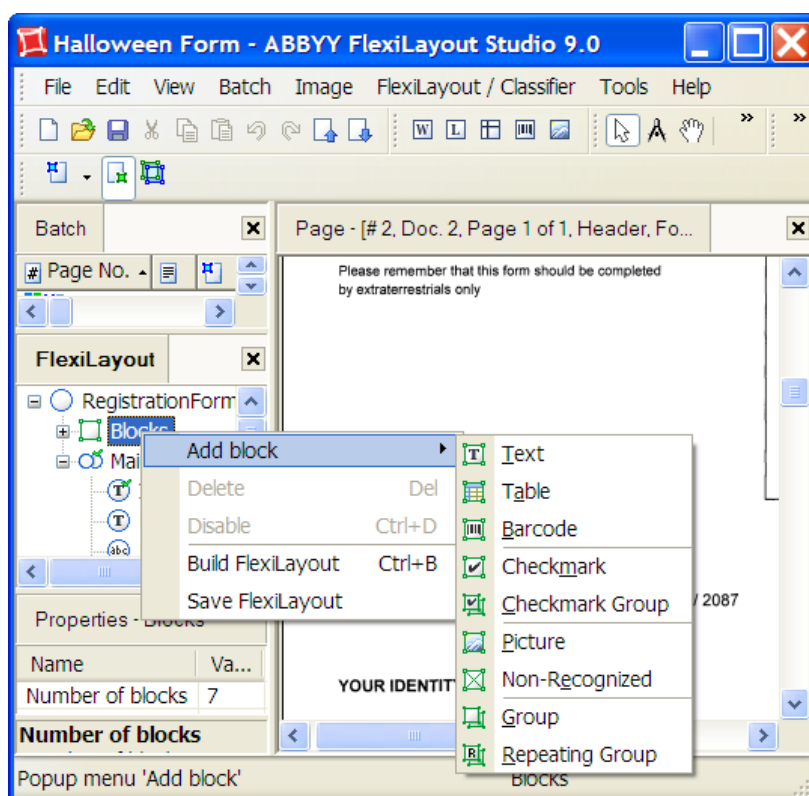
Once you have matched the FlexiLayout with the images, you will see that the **YOUR PLANET NAME** field corresponding to the **PlanetName** element has been found successfully on all the images.




Step 14. Describing the YOUR PLANET NAME field: PlanetName block

To create a block corresponding to the **YOUR PLANET NAME** field:

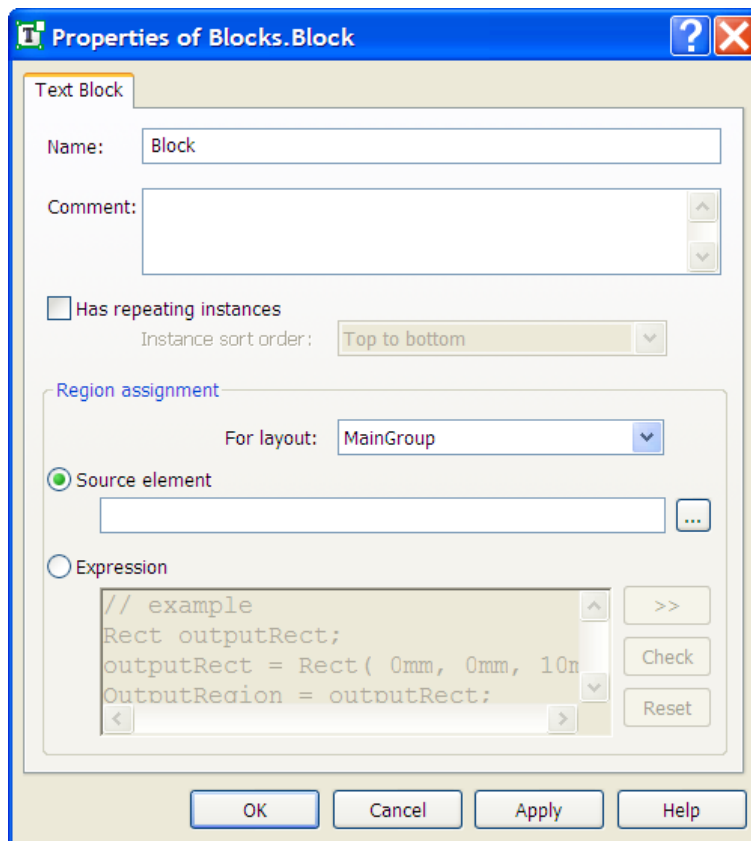
1. Click on the **FlexiLayout** tab in the program main window.
2. Select **Blocks** in the **FlexiLayout tree**.
3. Select **Add Blocks>Text** in the FlexiLayout menu or **New>Text** in the local menu.



4. In the **Properties** dialog box that opens, type a name for the block in the **Name** field, e.g. **PlanetName**

 **Note:** The name of the block need not coincide with the name of the element corresponding to the **YOUR PLANET NAME** field, but this is convenient when working with the FlexiLayout.

5. To describe the location of the block, select **Source element**. Click "... " and specify the **PlanetName** element as the source element.



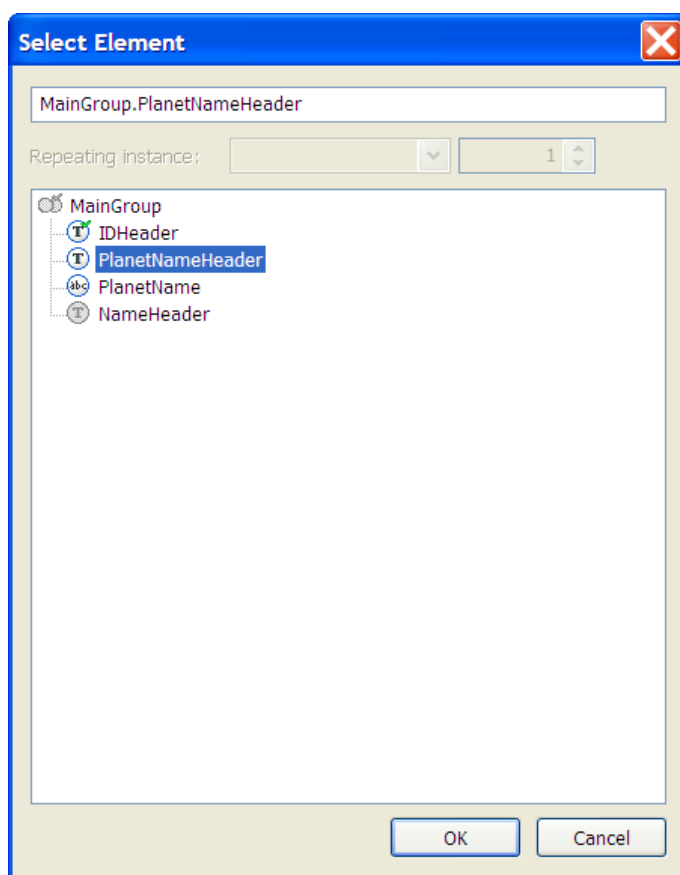
Step 15. Describing the NAME field

To describe the location of the block corresponding to the **NAME** field:

1. Create an element of the **Static Text** type and name it **NameHeader**. This element will correspond to the title of the **NAME** field.
2. Create an element of the **Character String** type and name it **GuestName**. This element will correspond to the **NAME** field proper.
3. Create a **GuestName** block corresponding to the **NAME** field.

The elements and the block for this field are created as described in Steps 10 to 14 with the following exceptions:

When specifying the search area for **NameHeader**, specify that it is located above the **PlanetNameHeader** element. This will prevent the program from formulating an erroneous hypothesis that contains the word **NAME** from the **YOUR PLANET NAME** title.



Unlike the alphabet of the **PlanetName** element, the alphabet of the **GuestName** element must include only **English** letters and the "_" symbol. In the **String Length** field, specify the fuzzy interval {0, 0, 50, INF }.

When specifying the search area for the **GuestName** element, specify the **NameHeader** element as the reference element.

When creating the **GuestName** block, specify the **GuestName** element as the source element.

Step 16. Describing the YOUR SPACESHIP NUMBER field

To describe the location of the block corresponding to the **YOUR SPACESHIP NUMBER** field:

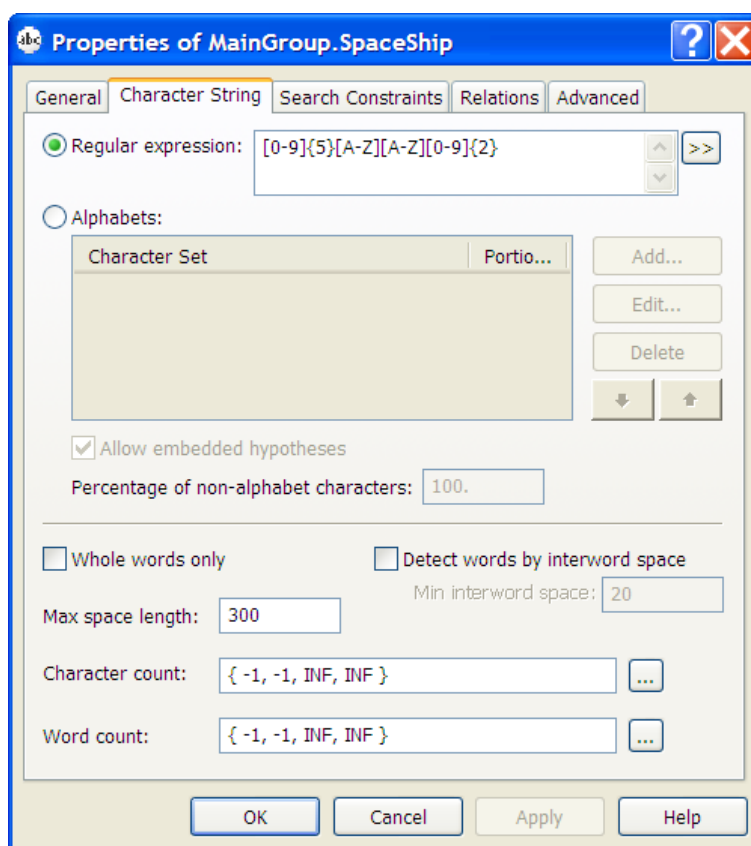
1. Create an element of the **Static Text** type and name it **SpaceShipHeader**. This element will correspond to the title of the **YOUR SPACESHIP NUMBER** field.
2. Create an element of the **Character String** type and name it **SpaceShip**. This element will correspond to the **YOUR SPACESHIP NUMBER** field proper.
3. Create a block and name it **SpaceShip**. This block will correspond to the **YOUR SPACESHIP NUMBER** field.

The elements and the block for this field are created as described in Steps 10 to 14 with the following exceptions:

Unlike for the **PlanetName** element, specify a regular expression for the **SpaceShip** element rather than an alphabet. This is necessary because the format of data in the **YOUR SPACESHIP NUMBER** field is the same on all the test images.

The format of spaceship numbers can be defined using the following regular expression:

[0-9]{5}[A-Z][A-Z][0-9]{2} (for more on regular expressions see the help file for ABBYY FlexiCapture)



When defining the search area for the **SpaceShip** element, specify the **SpaceShipHeader** element as the reference element.

When creating the **SpaceShip** block, specify the **SpaceShip** element as the source element.

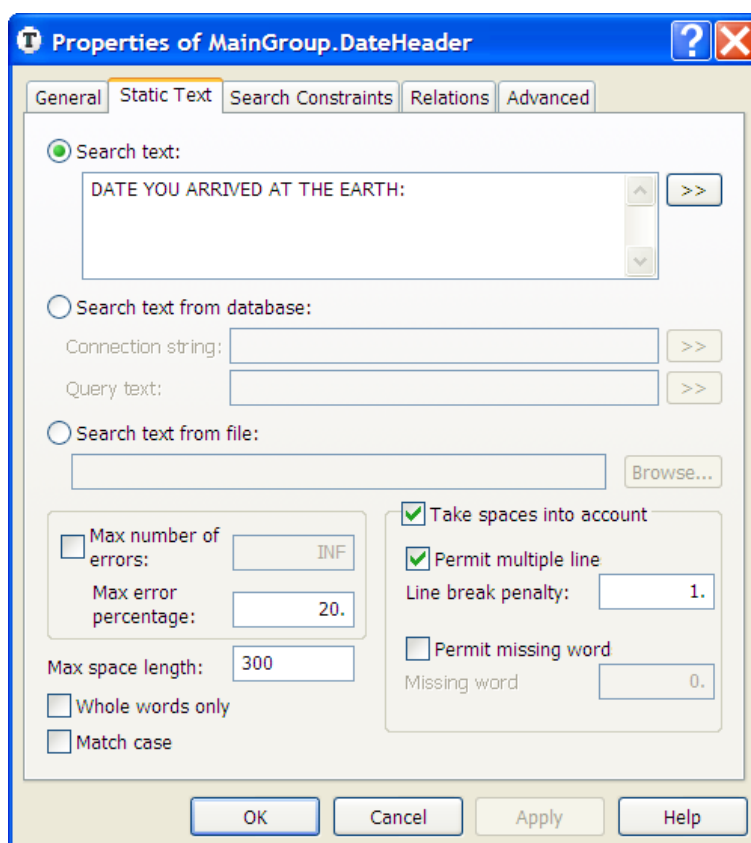
Step 17. Describing the DATE YOU ARRIVED AT THE EARTH field

To describe the location of the block corresponding to the **DATE YOU ARRIVED AT THE EARTH** field:

1. Create an element of the **Static Text** type and name it **DateHeader**. This element will correspond to the title of the **DATE YOU ARRIVED AT THE EARTH** field.
2. Create an element of the **Date** type and name it **ArrivalDate**. This element will correspond to the **DATE YOU ARRIVED AT THE EARTH** field proper.
3. Create a block and name it **ArrivalDate**. This block will correspond to the **DATE YOU ARRIVED AT THE EARTH** field.

The elements and the block for this field are created as described in Steps 10 to 14 with the following exceptions:

Unlike for the **PlanetNameHeader** element, enter the text in the **Search text** field with spaces: **DATE YOU ARRIVED AT THE EARTH**. Select the **Allow for spaces** and **Permit multiple lines** options and set **Line break penalty** to 1.



Properties of MainGroup.DateHeader

General Static Text Search Constraints Relations Advanced

☒ Search text:

DATE YOU ARRIVED AT THE EARTH: >>

☐ Search text from database:

Connection string: >>

Query text: >>

☐ Search text from file:

Browse...

☐ 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

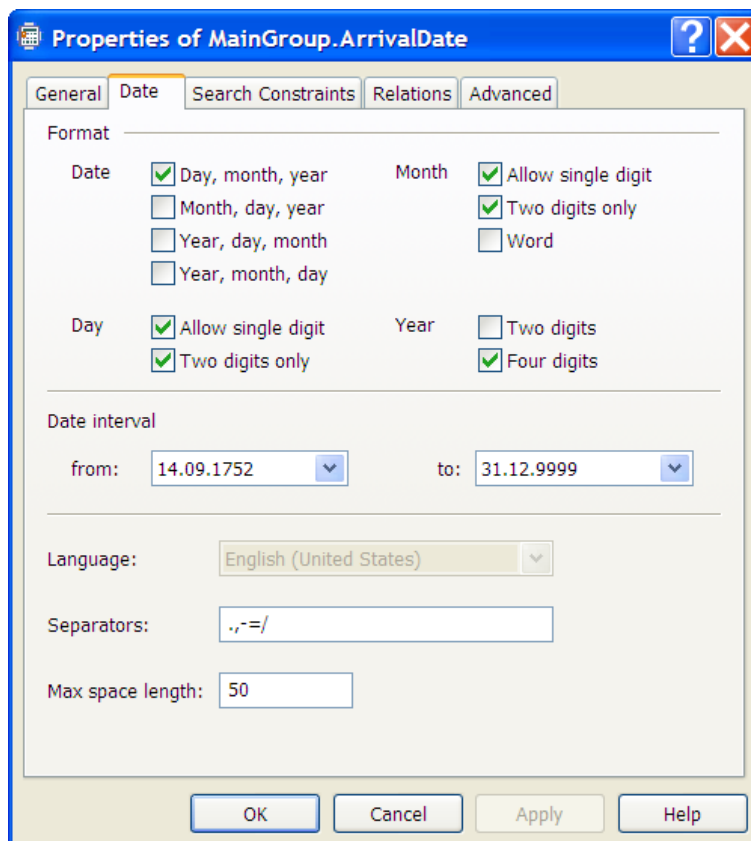
Line break penalty: 1.

☐ Permit missing word

Missing word: 0.

OK Cancel Apply Help

For the **DATE YOU ARRIVED AT THE EARTH** field, create an element of the **Date** type and name it **ArrivalDate**. For the **ArrivalDate** element, specify all the possible date formats on the **Date** tab. We assume that months will always be written in digits, and the year will always be a 4-digit number. Clear the **Word** box in the **Month** group and the **Two digits** box in the **Year** group. Other options may be left unchanged.



Properties of MainGroup.ArrivalDate

General Date Search Constraints Relations Advanced

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: 50

OK Cancel Apply Help

When defining the search area for the **ArrivalDate** element, specify the **DateHeader** element as the reference element.

When creating the **ArrivalDate** block, specify the **ArrivalDate** element as the source element.

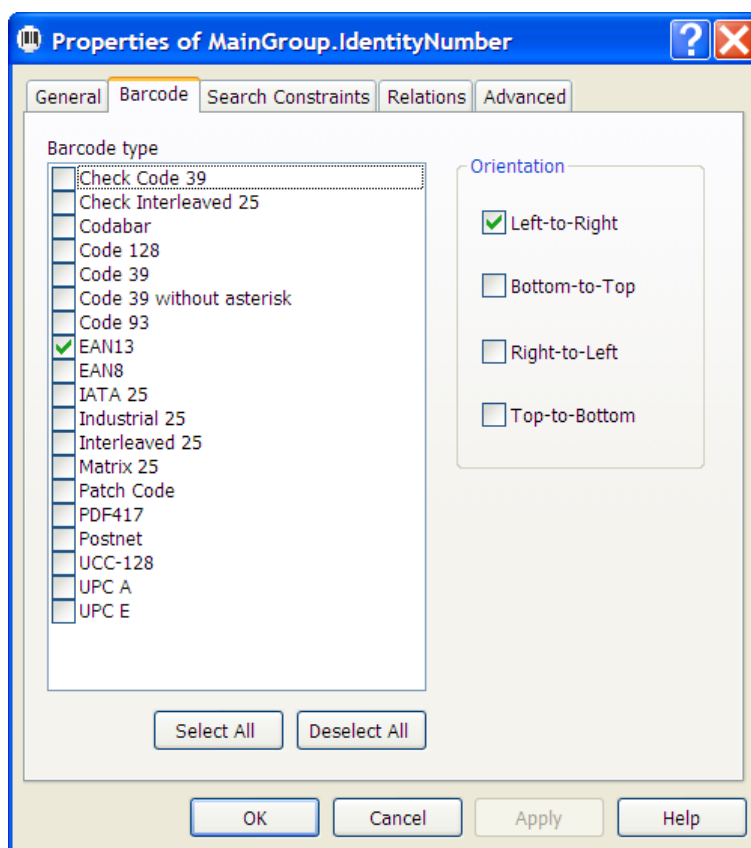
Step 18. Describing the YOUR IDENTITY NUMBER ON THE PARTY field

To describe the location of the block corresponding to the **YOUR IDENTITY NUMBER ON THE PARTY** field:

1. Create an element of the **Static Text** type and name it **BarcodeHeader**. This element will correspond to the title of the **YOUR IDENTITY NUMBER ON THE PARTY** field.
2. Create an element of the **Barcode** type and name it **IdentityNumber**. This element will correspond to the **YOUR IDENTITY NUMBER ON THE PARTY** field proper.
3. Create a block and name it **IdentityNumber**. This block will correspond to the **YOUR IDENTITY NUMBER ON THE PARTY** field.

The elements and block for this field are created as described in Steps 10 to 14 with the following exceptions:

For the **YOUR IDENTITY NUMBER ON THE PARTY** field, create an element of the **Barcode** type and name it **IdentityNumber**. For the **IdentityNumber** element, specify the type of barcode (**EAN13**) and its orientation (**Left-to-Right**) on the **Barcode** tab.




When creating an **IdentityNumber** block, specify the **IdentityNumber** element as the source element.

Step 19. Describing the ANY TEXT field

The field located at the bottom of the form has no fixed title. Therefore, objects other than titles must be used to locate this field. We will assume that each **ANY TEXT** field has its own unique title.

First, try using the already created **IdentityNumber** element to look for the **ANY TEXT** field, because on all the test images the **ANY TEXT** field is located below this element.

 **Note:** Do not use separators that are present on some images, as those are not regular and reliable elements.

To use the **IdentityNumber** element to find the object corresponding to the **ANY TEXT** field:

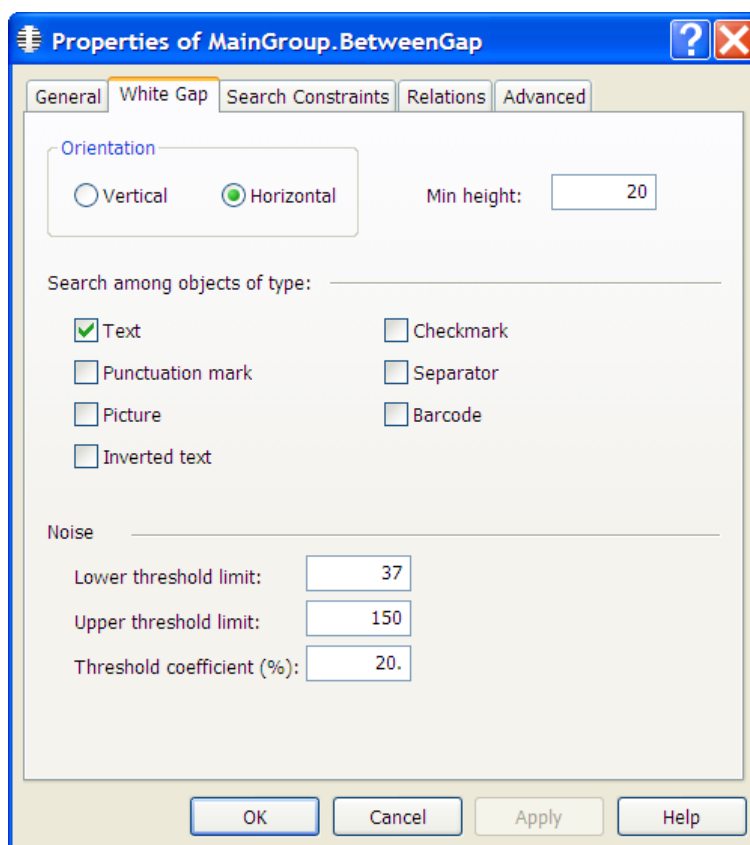
1. Create an element of the **Text Fragment** type and name it **TextField**. This element will correspond to the **ANY TEXT** field.
2. On the **Relations** tab, describe the location of the **TextField** element as **Below** the **IdentityNumber** element.

Try finding the **ANY TEXT** field using the **IdentityNumber** element. You will see that on some images the hypothesis will also include barcode digits. Therefore, you need to describe the location of the **ANY TEXT** field more precisely.

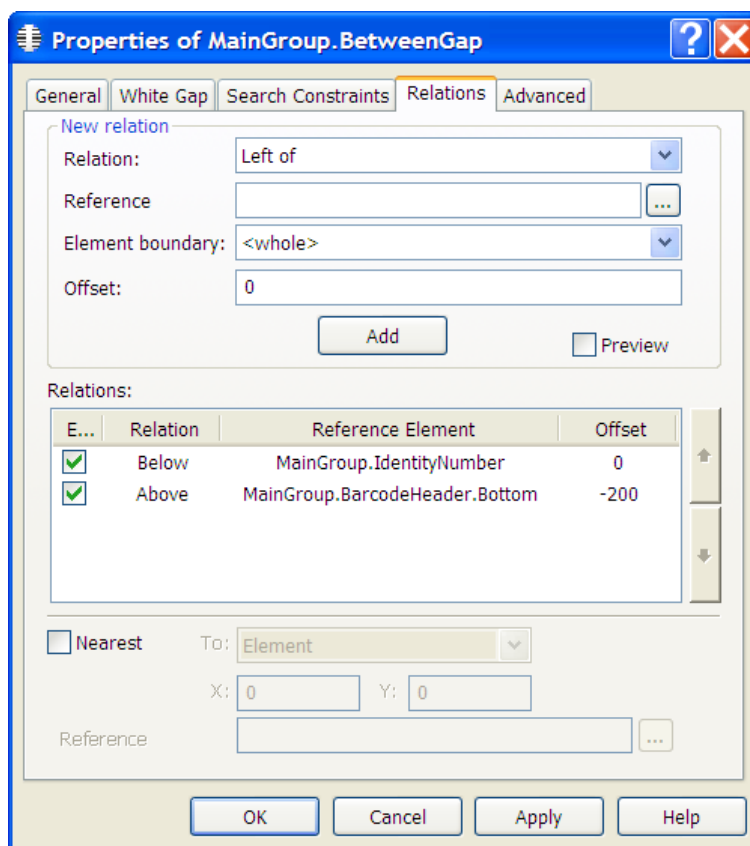
Create a **White Gap** element with horizontal orientation. This element will correspond to the gap between the text and the barcode. This element will be used as a reference element when looking for the **ANY TEXT** field.

Be sure to place the **White Gap** element above the **TextField** element in the elements tree!

1. Create an element of the **White Gap** type and name it **BetweenGap**.
2. On the **White Gap** tab, select **Horizontal** as the orientation of the **BetweenGap**.



3. On the **Relations** tab, specify the search area for the **BetweenGap** element as an area below the **IdentityNumber** element but no further than 200 dots from the bottom edge of the **BarcodeHeader** element.



4. Create an element of the **Text Fragment** type and name it **TextField**.
5. On the **Relations** tab, describe the location of the **TextField** element as **Below** the **BetweenGap** element, **Offset=0**.

Try matching the FlexiLayout with all the test images. You will see that the formulated hypothesis about the **TextField** element does not contain any unwanted data. This means that the **BetweenGap** element can be used to detect the **TextField** element.

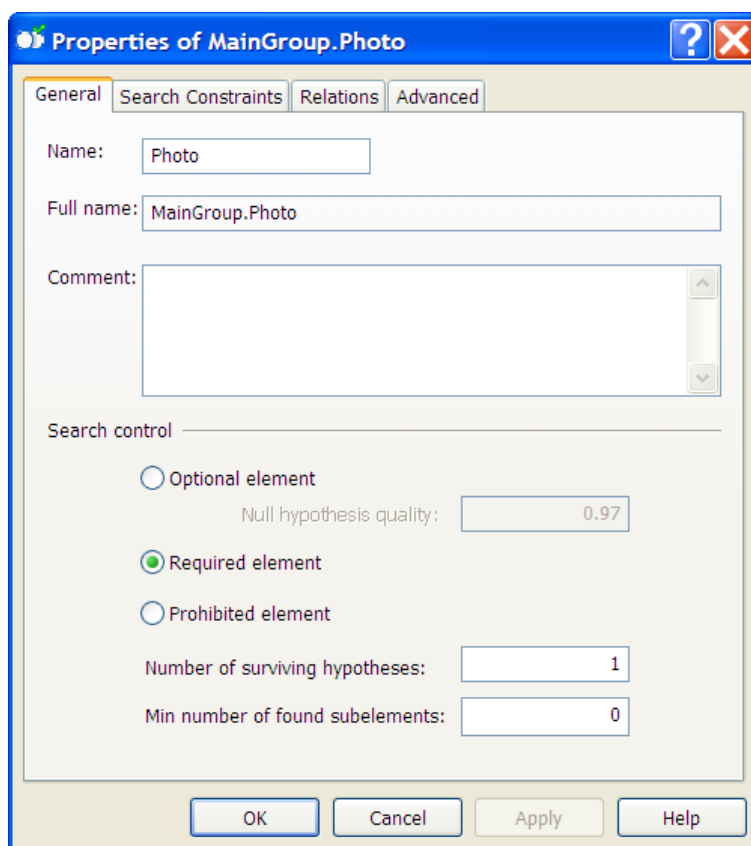
To describe the location of the block corresponding to the **ANY TEXT** field, create a block of the **Text** type, name it **TextField** and specify the **TextField** element as the source element.

Step 20. Describing the YOUR PHOTO IN FANCY DRESS field

As has been noted in Step 6, not all the parts of the photo on images 1 and 6 were recognized as the **Image Object** of the **Picture** type.

To get a full photo, the program needs to find all the types of **Image Object** located in the photo area.

For the sake of convenience, group all the photo elements in one compound **Photo** element. This will also limit the number of formulated hypotheses.



The compound **Photo** element (of the **Group** type) will include:

- a **PhotoPicture** element of the **Object Collection** type corresponding to the photo;
- a **PhotoHeader** element of the **Static Text** type corresponding to the title **YOUR PHOTO IN FANCY DRESS**;
- four separators around the photo which are required for finding the **PhotoPicture** element.

As the text of the title (namely, **YOUR PHOTO IN FANCY DRESS**) occurs nowhere else in this document, we added the corresponding element first and will search for the separators relative to this element.

To create an element corresponding to the field title:

1. Create an element of the **Static Text** type and name it **PhotoHeader**.
2. The properties of the element are set similarly to the other field titles as described above.

Try matching the FlexiLayout with the images. You will see that the program reliably detects the title on all the images.

Create **Separator** elements for the four separators around the photo and give them the following names:

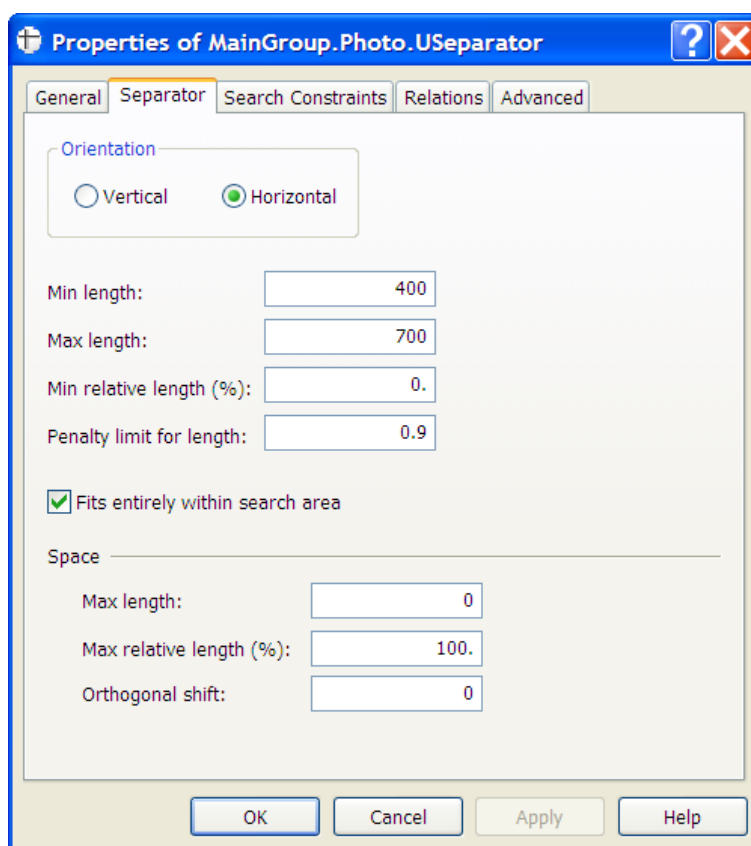
USeparator – upper separator

BSeparator – lower separator

LSeparator – left separator

RSeparator – right separator

On the **Separator** tab in the **Properties** dialog box, specify the orientation of each of the 4 separators (**Vertical** or **Horizontal**). Do not change the other properties.



Now use the test images to specify the search area for each of the separators.

Upper separator (**USeparator** element):

below the photo caption but not farther than 100 dots from its top border (**Below** the **PhotoHeader** element and **Above** the **PhotoHeader**, **Offset** = -100, **Element Border** = top).
 within the area which is not longer than the length of the caption by more than 80 dots on each side (**Leftof** the right border of the **PhotoHeader** element, **Offset** = -80, **Element Border** = right and **Rightof** the left border of the **PhotoHeader** element, **Offset** = -80, **Element Border** = left).

Lower separator (**BSeparator** element):

below the title of the NAME field (**Below** the **NameHeader** element),
 above bottom of **IdentityNumber** (**Above** the **IdentityNumber** element, **Element Border** = bottom).
 within the area which is not longer than the length of the title by more than 80 dots on each side (**Leftof** the right border of the **PhotoHeader** element, **Offset** = -80, **Element Border** = right and **Rightof** the left border of the **PhotoHeader** element, **Offset** = -80, **Element Border** = left).

Left separator (**LSeparator** element):

within the area which is not wider than the distance between the upper and lower separators by more than 50 dots on each side (**Below** the **USeparator** element, **Offset** = -50, and **Above** the **BSeparator** element, **Offset** = -50).
 the separator is located not far from the left border of the upper separator, the distance cannot exceed 40 dots (**Rightof** the left border of the **USeparator** element, **Offset** = -20, **Element Border** = left, and **Leftof** the left border of the **USeparator** element, **Offset** = -20, **Element Border** = left).

Right separator (**RSeparator** element):

within the area which is not wider than the distance between the upper and lower separators by more than 50 dots on each side (**Below** the **USeparator** element, **Offset** = -50, and **Above** the **BSeparator** element, **Offset** = -50).
 the separator is located not far from the right edge of the upper separator, the distance cannot exceed 40 dots (**Rightof** the right border of the **USeparator** element, **Offset** = -20, **Element Border** = right, and **Leftof** the right border of the **USeparator** element, **Offset** = -20, **Element Border** = right).

Now if you try matching the FlexiLayout with the images you will see that the program can detect these areas. In the **Tree of Hypotheses**, right-click on a hypothesis and select **Show Search Area** in the local menu. If the search criteria are wrong and the program failed to locate the search area properly, you will immediately notice this. You may also see that even the search area of a correct hypothesis needs some adjustment, e.g. it may be too small.

The maximum size of all the separators was detected by trial and error while studying the properties of Separator objects during pre-recognition.

To describe the **YOUR PHOTO IN FANCY DRESS** field:

1. Create an element of the **Object Collection** type and name it **PhotoPicture**.
2. On the **Relations** tab, specify the search area for the **PhotoPicture** element:

below the upper separator (**Below** the **USeparator** element)
 above the lower separator (**Above** the **BSeparator** element)
 to the right of the left separator (**Righthof** the **LSeparator** element)
 to the left of the right separator (**Leftof** the **RSeparator** element)

Then create the **PhotoPicture** block and specify the **PhotoPicture** element as the source element.

Step 21. Exporting the FlexiLayout

Before you can use your FlexiLayout in ABBYY FlexiCapture, you must export it into AFL format:

1. Select the Export **FlexiLayout...** command in the File menu.
2. In the **Export FlexiLayout** dialog box that opens, enter a name for the file and select the folder in which to save the file.

The resulting *.afl file can then be used in ABBYY FlexiCapture.


Step 22. Opening the FlexiLayout in ABBYY FlexiCapture

The FlexiLayout saved in *.afl format can be opened in ABBYY FlexiCapture:

1. Launch ABBYY FlexiCapture.
2. Create a new project or open an existing project.
3. Create a new document template or open an existing template in the Template Editor.

 **Note:** to open an existing template, go to **Project>Document Templates...**, select the required template from the list of available templates and click **Edit...**

4. In the **Template Editor**, open the **Section Properties** dialog box (**View>Properties...**).
5. Click on the **Flexible Description** tab.
6. Click the **Load...** button.
7. In the **Open** dialog box that opens, select the required *.afl file.

 **Note:** The required template blocks will be created automatically based on the FlexiLayout. To view the blocks, select the **Match Section** command in the **Tools** menu.

Halloween Form* - Document Template Editor

Template Edit View Document Tools Help

Check

NOTE:
Please remember that this form should be completed by extraterrestrials only

NAME: MARTIAN THE FIRST

YOUR PLANET NAME: MARS (satellite Phobos)

YOUR SPACESHIP NUMBER: 123 45 QU 16

DATE YOU ARRIVED AT THE EARTH: 12/12/2099

YOUR IDENTITY NUMBER ON THE PARTY IS:

YOUR PHOTO IN FANCY I

1 2 3 4 5 6 7 8 9 9 9 9

Document Struct...

Fields Pages

Document Section1

- PlanetName
- GuestName
- SpaceShip
- ArrivalDate
- IdentityNumber
- TextField
- PhotoPicture

75%

Cuts part from selected region

Document Section1: PlanetName Text