User's Guide

Script Extractor

SCREEN SYSTEMS

Printed on 18 May, 2017

Contents

Overview of Script Extractor 3

What does Script Extractor do? 4
About scripts....5
Extracting content from the script....6
About models 7

The Script Extractor screen 9

The List pane....10

Getting started 13

Loading a script 13
Topping and tailing the script 14
Training Script Extractor....15
Extracting the required content 16
Fine-tuning the extraction 18
Saving the results....19

Assessing extracted content 21

Synchronising script and list items 21
Showing or hiding columns 22
Sorting by column....23
Filtering results 24
Finding a specific word or phrase 25

Techniques for refining results 27

Resetting the classification....27
Manually overriding the classification....27
Undoing a classification 29
Finding and replacing text 30
Deciding the most effective approach 32

Using models 33

Creating a new model 33
Using a saved model....34
Updating an existing model....34
Clearing a model 35
Working in batch mode....36

Configuring Script Extractor 37

The configuration options 37

Troubleshooting and tips 39

Troubleshooting....39
Tips....40

The Script Extractor toolbar and menus 43

Overview of Script Extractor

Script Extractor helps you to quickly extract content useful for subtitling or translation from production scripts, and can significantly speed up your workflow.

With Script Extractor, you identify one or two examples of each type of content you are interested in (for example, Speaker label, Dialogue, Timecode, Scene Boundary). The system then recognizes and picks up all similar items from the script based on their layout and style. If the script is consistently laid out, you may find that you can save the results straight away, otherwise you can quickly adjust the classifications if needed. Either way, the process of extracting useful content is much quicker than having to process the script manually.

Script Extractor requires Microsoft Word to be available on your computer.

What does Script Extractor do?

Script Extractor recognises different kinds of script content based on examples you provide, and then categorises the content 'elements' into different groups or 'classes'. You can then extract and save only those you are interested in. For example, you may want to extract timecodes, dialogue and speaker identities, but ignore camera directions or actor directions.

Depending on the script layout, Script Extractor may also be able to identify and extract scene boundaries or translation notes among the elements you require.

Script Extractor's ability to classify script elements depends on the style and layout of the script - see: ***About scripts*** (page 5). In some instances, you may need to manually fine-tune the classifications to ensure that script elements are identified correctly and can be correctly extracted.

Often, the extraction produces just the items you require from the script based only on your initial examples. With other scripts, the results may be close, but not entirely accurate (for example, unwanted items may be included) and in this situation you will need to adjust the results manually.

Script Extractor classifications

Script Extractor classifies each item in the script into one of the following categories:

- Speaker ID - a label which indicates who is speaking
- Dialogue - the spoken words
- Timecode - In or Out points (in normal timecode format)
- Boundary - the start of a scene
- Production - production directions
- Unwanted - for the purposes of extraction, these items are ignored

These are the core classifications, though with some implementations further classifications may be available.

For more information on extracting the information you require, see: ***Extracting script content*** (page 6).

About scripts

Script Extractor is able to process Microsoft Word files (.doc and .docx files) and Microsoft rich text format (.rtf files). Scripts in any other format, for example .pdf, must first be converted into one of the suitable file formats.

Script Extractor is most efficient when the script being processed has a reasonably consistent layout, that is:

- each type of content has a distinct position (indented or in table columns) or
- each type of content is presented in a distinct style (such as bold or underlined).

Consistency in layout and style means that Script Extractor can reliably identify items that are similar to your examples and assign them the same classification.

The example script below is mainly consistent, so once Script Extractor has determined the layout and positioning used to identify the speaker labels, for example, it will be able to categorise elements that are similarly presented as also being the speaker ID.

SCENE 1: BEACH, OUTSIDE, DAY

<WAVES LAP AGAINST THE SHORE ON A TROPICAL ISLAND>

00:00:10:10 NARRATOR The average person can hold their breath for 40 seconds.

<A WOMAN IN A BIKINI DIVES INTO THE WATER>

00:00:15:13 NARRATER Tanya Streeter can do it for over six minutes.

<SHE GRACEFULLY SWIMS OVER TO THE CAMERA>

00:00:17:12 CAPTION Free Diving.

00:00:19:17 TANYA It's just my thing (sighs).

00:00:20:21 I have this unique relationship with the ocean.

00:00:24:00 I'm entirely blessed to be able to test myself in that environment

00:00:30:04 NARRATOR Tanya uses this ability to dive to great depths without equipment,

00:00:34:08 using only the air in her lungs.

SCENE 2: UNDERWATER

<CUT TO LIBRARY FILM OF TANYA DURING A DIVE>

However, if your script has different information types presented in the same style/position, or a given information type is presented in different ways, Script Extractor may not be able to reliably distinguish one type from the other. In the script below, Script Extractor may classify the scene descriptions and the speaker IDs as being the same.

SCENE 1: BEACH, OUTSIDE, DAY

WAVES LAP AGAINST THE SHORE ON A TROPICAL ISLAND

00:00:10:10	NARRATOR	The average person can hold their breath for 40 seconds.

A WOMAN IN A BIKINI DIVES INTO THE WATER

00:00:15:13	NARRATER	Tanya Streeter can do it for over six minutes.

SHE GRACEFULLY SWIMS OVER TO THE CAMERA

00:00:17:12	CAPTION	Free Diving.
00:00:19:17	TANYA	It's just my thing (sighs).
00:00:20:21		I have this unique relationship with the ocean.
00:00:24:00		I'm entirely blessed to be able to test myself in that environment.
00:00:30:04	NARRATOR	Tanya uses this ability to dive to great depths without equipment,
00:00:34:08		using only the air in her lungs.

When a script is inconsistent - for example, where different elements have a similar layout, or where a given type of content is presented in different ways - you may need to fine-tune the classifications manually so as to ensure that all the required information is extracted.

Extracting content from the script

Generally, there are 6 steps to extracting the content you require from a script.

1. Load the script into Script Extractor and review the script elements
2. Train Script Extractor to recognise the elements you wish to extract
3. Use Script Extractor to isolate the content you are interested in
4. Review the results
5. Fine-tune by:
   - adjusting the classification manually
   - providing more examples
   - overriding the classifications of given items

   and then extract the content again
6. Save the results. The saved file can then be imported for subtitling or translation workflows.

These steps are described more fully in ***Getting started*** (page 13).

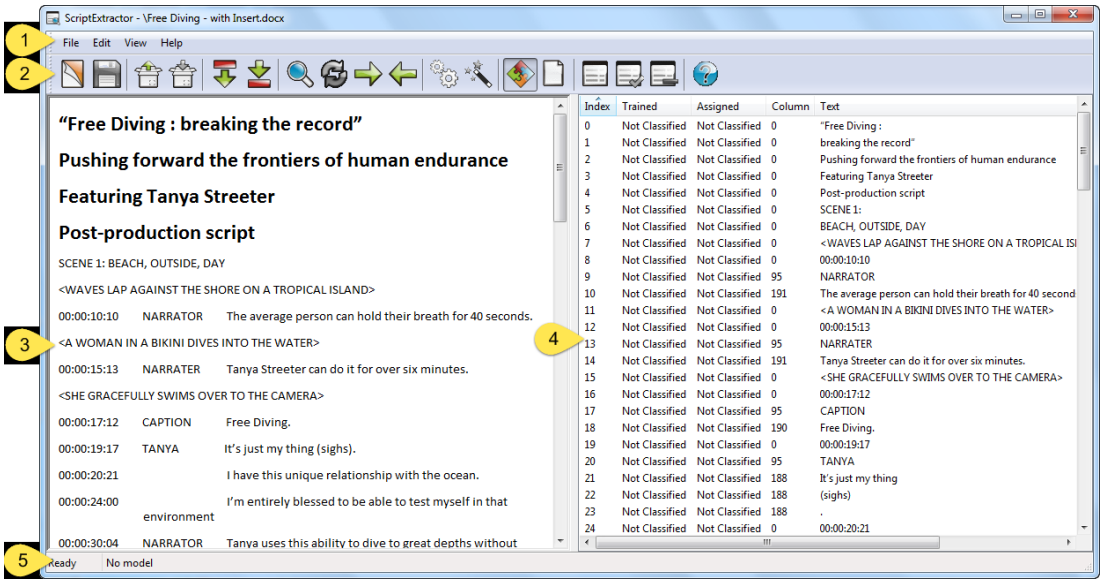
About models

If you process a lot of scripts that have the same layout (for example, multiple scripts from the same producer) you can save time by preserving your example classifications for re-use next time you load a script of that type. You do this by creating a *model*.

Creating a model is an optional step once you have extracted the script content you require. See: ***Using models*** (page 33) for more information.

The Script Extractor screen

When you launch Script Extractor, the main screen will look similar to the annotated screen shown below.



Reference	Function
1	**Main menu** - most functions are available from this menu hierarchy
2	**Toolbar** - contains the most frequently used functions. Some of the toolbar options are active only when a script is loaded
3	**Script pane** - this pane shows a preview of the script. When you start Script Extractor this area will be blank until you load a script
4	**List pane** - depending on where you are in the workflow, this pane displays one of the following ▪ a list of the script elements and additional information about them, such as their classification, column position and typeface, as is shown above ▪ the text of just the items that have been extracted from the script See: ***The List pane*** (page 10) for more information.
5	**Status bar**

For more information about the functions on the main menu and the toolbar options, see: ***The Script Extractor toolbar and menus*** (page 43).

The List pane

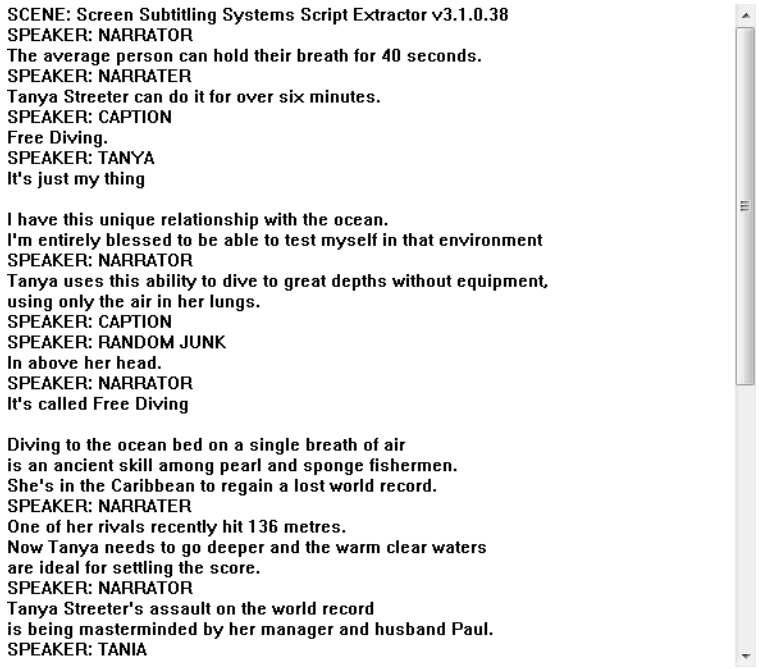
The content of the List pane depends on whether you choose to display

- The result of the last extraction (called Show Dialogue), or
- The list of script elements and their classification (Show Classifications).

See: ***The Script Extractor toolbar and menus*** (page 43) for more information on the controls for the List pane display.

Show Dialogue

When Show Dialogue is selected, the List pane displays the extraction results in a simple text file format, as in the example below. The data types that are included depend on how you tailor the extraction.

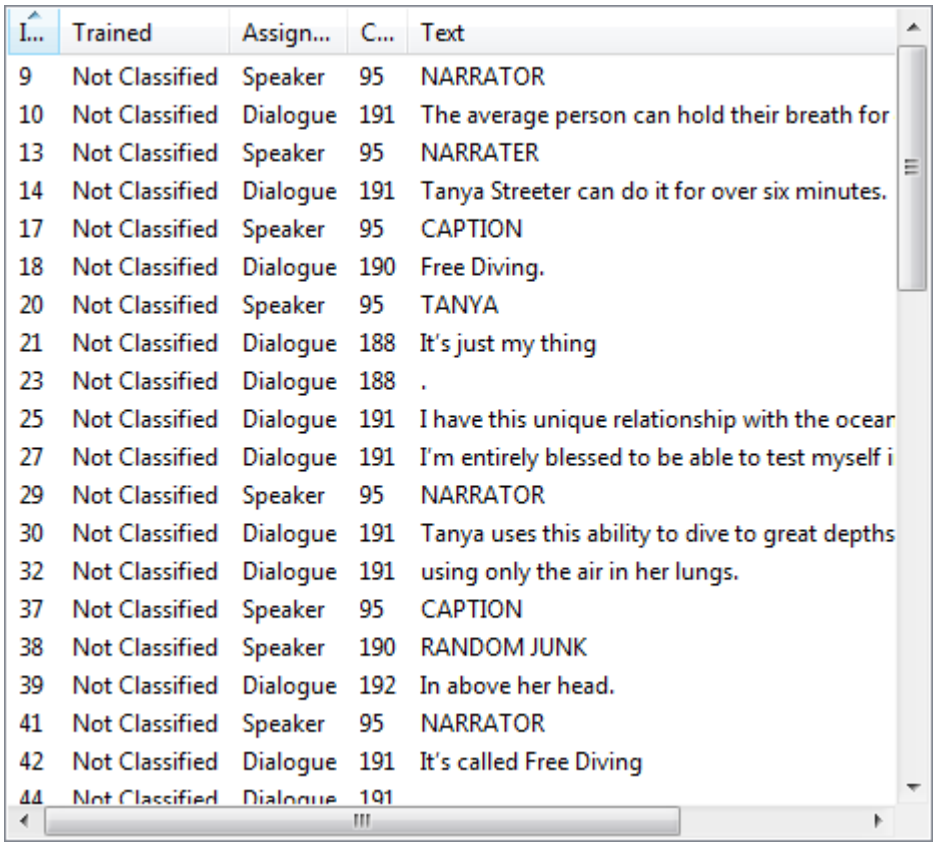


Show Classifications

When Show Classifications is selected, the List pane displays script elements and additional information about them. You can choose to display

- All the elements in the script
- Only those elements from the script *included* in the last extraction
- Only those elements from the script *omitted* from the last extraction.

The second of these shows the same content items as Show Dialogue, it is just formatted differently, as shown in the example below.



I...	Trained	Assign...	C...	Text
9	Not Classified	Speaker	95	NARRATOR
10	Not Classified	Dialogue	191	The average person can hold their breath for
13	Not Classified	Speaker	95	NARRATER
14	Not Classified	Dialogue	191	Tanya Streeter can do it for over six minutes.
17	Not Classified	Speaker	95	CAPTION
18	Not Classified	Dialogue	190	Free Diving.
20	Not Classified	Speaker	95	TANYA
21	Not Classified	Dialogue	188	It's just my thing
23	Not Classified	Dialogue	188	.
25	Not Classified	Dialogue	191	I have this unique relationship with the ocear
27	Not Classified	Dialogue	191	I'm entirely blessed to be able to test myself i
29	Not Classified	Speaker	95	NARRATOR
30	Not Classified	Dialogue	191	Tanya uses this ability to dive to great depths
32	Not Classified	Dialogue	191	using only the air in her lungs.
37	Not Classified	Speaker	95	CAPTION
38	Not Classified	Speaker	190	RANDOM JUNK
39	Not Classified	Dialogue	192	In above her head.
41	Not Classified	Speaker	95	NARRATOR
42	Not Classified	Dialogue	191	It's called Free Diving
44	Not Classified	Dialogue	191	

Getting started

The topics in this section take you through the process of extracting content from the script. Often, the most efficient way to use the system is to have a quick trial-run on your script. If this first run-through does not return good enough results, you can then decide whether more training examples and/or some manual 'tidying up' would be the best way to obtain only the content that you require.

Loading a script

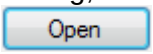
To load a script

1. Ensure the script to be processed is in a suitable format (Microsoft Word or rich text format).

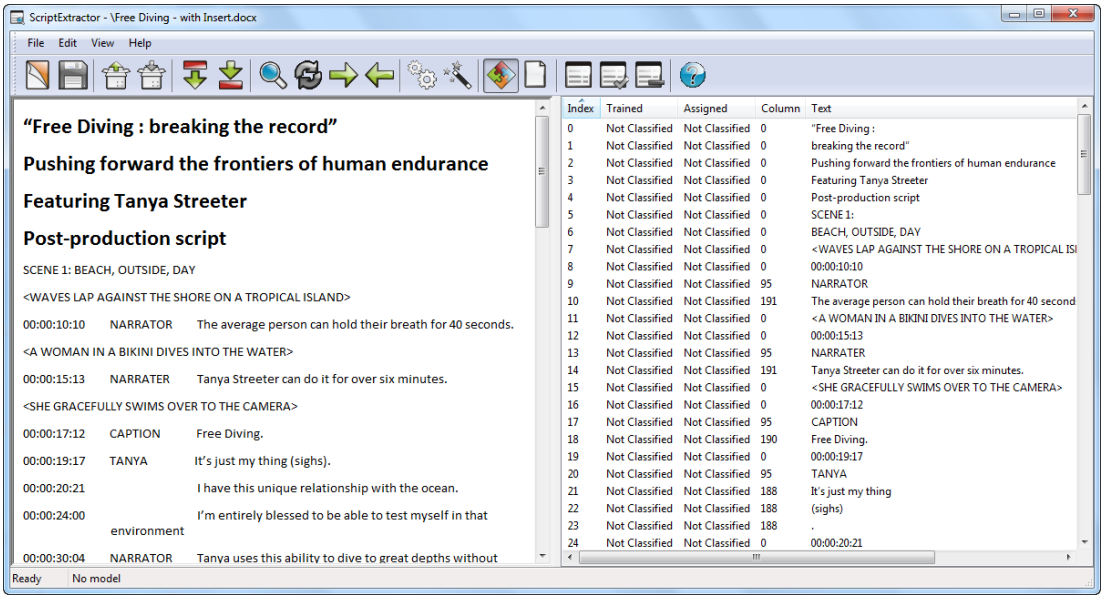
   If the file is not in a suitable format, you will first need to convert the script into one of the above formats.

2. Launch Script Extractor and do one of the following.
   - From the main menu, select **File>Open**.
   - Press **Ctrl+O**.
   - Click in the toolbar.

   This launches the **Open** dialog.

3. Within the **Open** dialog, find and select the file containing the script you wish to load, then click  .

   After a few moments you will see the script displayed in the Script pane, and then the List pane will display the script content elements with some additional information, similar to the screenshot below.



4 If the script contains header or footer information that is not strictly script content, you can ***top and tail*** (page 14) the script so that this information can be ignored by Script Extractor.

Once you have loaded the script you are ready to review the content and start ***training Script Extractor*** (page 15).

Topping and tailing the script

Topping and tailing immediately after you have loaded a script allows you to “grey out” any material at the start and/or end that is not dialogue-related. Such material may include production company logos, credits, actor lists or synopses. When Script Extractor processes the script, it ignores any content that you have greyed out and such content is always shown as 'Not classified'.

To top and tail the script

1 If there is content to be greyed out at the start of the script, highlight the first line of the useful dialogue content in the List pane.

You can do this either by clicking on that item within the List pane, or by ***finding it from the script*** (page 21).

2 Do one of the following to set the in-point for Script Extractor.

- Click in the toolbar.
- Press **Ctrl+B**.
- From the main menu, select **Edit>Mark Begin**.

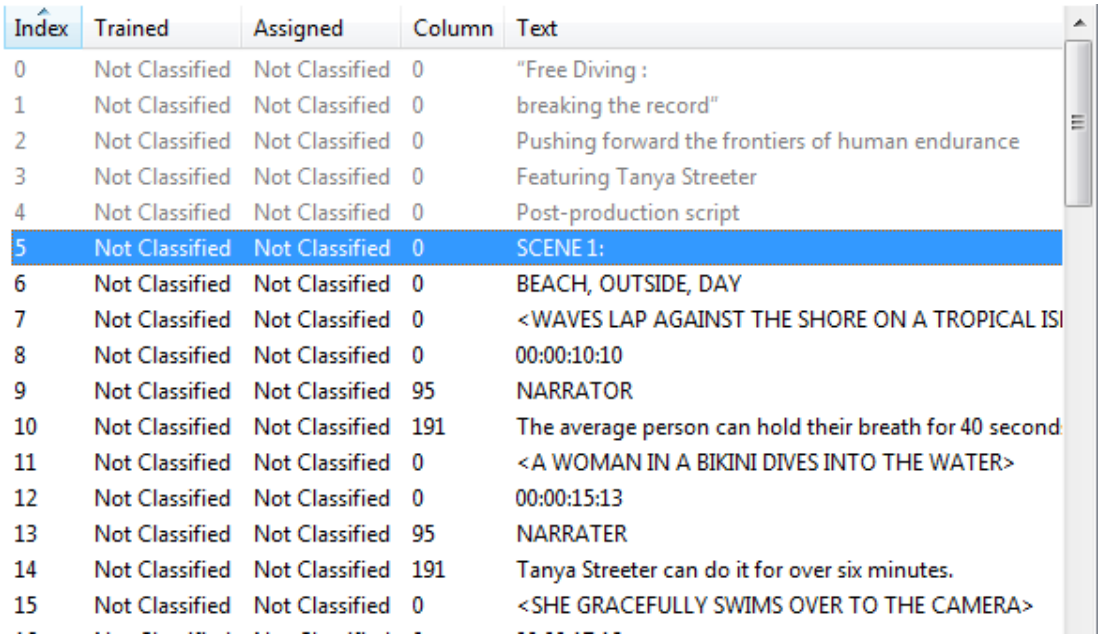
Any text in the script above the in-point that you have set will be displayed greyed out in the List pane.

3 If the script includes footer material that you wish to ignore, highlight the last line of the script proper (either by clicking on that item within the List pane, or by finding it from the script), then continue as described below.

4 Do one of the following to set the out-point for Script Extractor.

- Click in the toolbar.
- Press **Ctrl+E**.
- From the main menu, select **Edit>Mark End**.

The screenshot below shows the result of loading and marking the in-point of a script.



Index	Trained	Assigned	Column	Text
0	Not Classified	Not Classified	0	"Free Diving :
1	Not Classified	Not Classified	0	breaking the record"
2	Not Classified	Not Classified	0	Pushing forward the frontiers of human endurance
3	Not Classified	Not Classified	0	Featuring Tanya Streeter
4	Not Classified	Not Classified	0	Post-production script
5	Not Classified	Not Classified	0	SCENE 1:
6	Not Classified	Not Classified	0	BEACH, OUTSIDE, DAY
7	Not Classified	Not Classified	0	<WAVES LAP AGAINST THE SHORE ON A TROPICAL ISI
8	Not Classified	Not Classified	0	00:00:10:10
9	Not Classified	Not Classified	95	NARRATOR
10	Not Classified	Not Classified	191	The average person can hold their breath for 40 second
11	Not Classified	Not Classified	0	<A WOMAN IN A BIKINI DIVES INTO THE WATER>
12	Not Classified	Not Classified	0	00:00:15:13
13	Not Classified	Not Classified	95	NARRATER
14	Not Classified	Not Classified	191	Tanya Streeter can do it for over six minutes.
15	Not Classified	Not Classified	0	<SHE GRACEFULLY SWIMS OVER TO THE CAMERA>

Training Script Extractor

Training involves identifying typical examples of the content you wish to extract and classifying them as Speaker, Dialogue, Timecode, Boundary or Production items. You may also optionally label items as Unwanted - any such items and any like them will be ignored. You will find a reminder of the classifications under the Help menu.

Note: Unless the elements you wish to ignore are obviously unwanted and very distinct from other script elements, it may be preferable not to use 'Unwanted' for the initial extraction.

One or two examples of each type are generally enough to train Script Extractor, which uses these examples to decide a classification for other items in the script.

When you first load a script, all of the items shown in the List pane are marked as 'Not Classified' in both the 'Trained' and 'Assigned' columns. When you label an item, its status in the 'Trained' column changes to show the classification you have given it. When Script Extractor processes the script, the classification it assigns is shown in the 'Assigned' column.

To identify examples of content

1 In the List pane, use any of the following:
   - the vertical scroll bar
   - **PageUp** and **PageDown** keys
   - keyboard up and down arrow keys

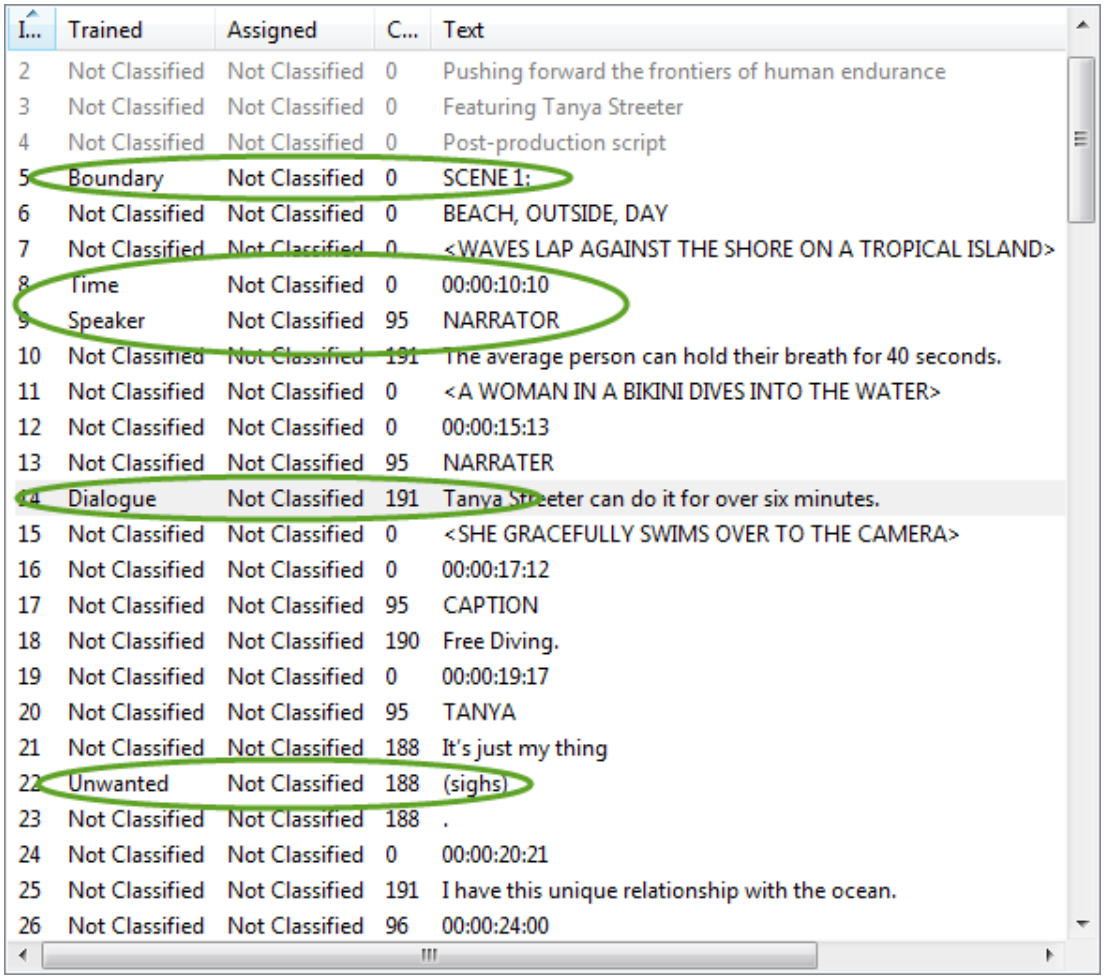
   to find an item you wish to label, then click it to select it.
2 When the item is selected, press one of the following keys to label it.
   - S - Speaker ID
   - D - Dialogue

- T - Timecode
- B - Boundary
- P - Production
- U - Unwanted

3 Find a couple of typical but different examples of each type of content you may be interested in and label them as described above.

For example, with speaker labels, choose different speakers. When choosing dialogue, select a long utterance and a short one.

4 Ensure that you have classified at least two items of each of the kinds of information you are interested in.



I...	Trained	Assigned	C...	Text
2	Not Classified	Not Classified	0	Pushing forward the frontiers of human endurance
3	Not Classified	Not Classified	0	Featuring Tanya Streeter
4	Not Classified	Not Classified	0	Post-production script
5	Boundary	Not Classified	0	SCENE 1:
6	Not Classified	Not Classified	0	BEACH, OUTSIDE, DAY
7	Not Classified	Not Classified	0	<WAVES LAP AGAINST THE SHORE ON A TROPICAL ISLAND>
8	Time	Not Classified	0	00:00:10:10
9	Speaker	Not Classified	95	NARRATOR
10	Not Classified	Not Classified	191	The average person can hold their breath for 40 seconds.
11	Not Classified	Not Classified	0	<A WOMAN IN A BIKINI DIVES INTO THE WATER>
12	Not Classified	Not Classified	0	00:00:15:13
13	Not Classified	Not Classified	95	NARRATER
14	Dialogue	Not Classified	191	Tanya Streeter can do it for over six minutes.
15	Not Classified	Not Classified	0	<SHE GRACEFULLY SWIMS OVER TO THE CAMERA>
16	Not Classified	Not Classified	0	00:00:17:12
17	Not Classified	Not Classified	95	CAPTION
18	Not Classified	Not Classified	190	Free Diving.
19	Not Classified	Not Classified	0	00:00:19:17
20	Not Classified	Not Classified	95	TANYA
21	Not Classified	Not Classified	188	It's just my thing
22	Unwanted	Not Classified	188	(sighs)
23	Not Classified	Not Classified	188	.
24	Not Classified	Not Classified	0	00:00:20:21
25	Not Classified	Not Classified	191	I have this unique relationship with the ocean.
26	Not Classified	Not Classified	96	00:00:24:00

Once you have entered your training examples you are ready to ***extract content*** (page 16).

Extracting the required content

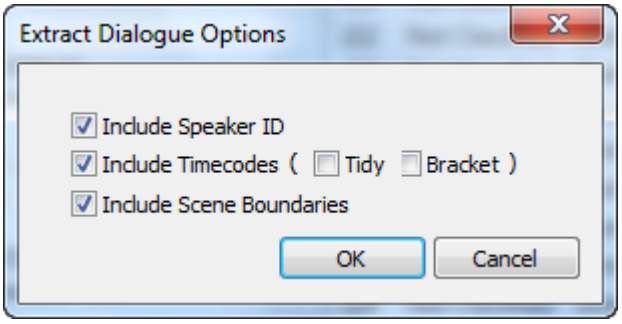
To extract content

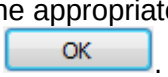
1 Ensure you have provided a few ***training*** (page 15) examples within the List pane.

2 Do one of the following.

- Click in the toolbar.
- From the main menu, select **File>Extract Dialogue**.
- Press **F7**.

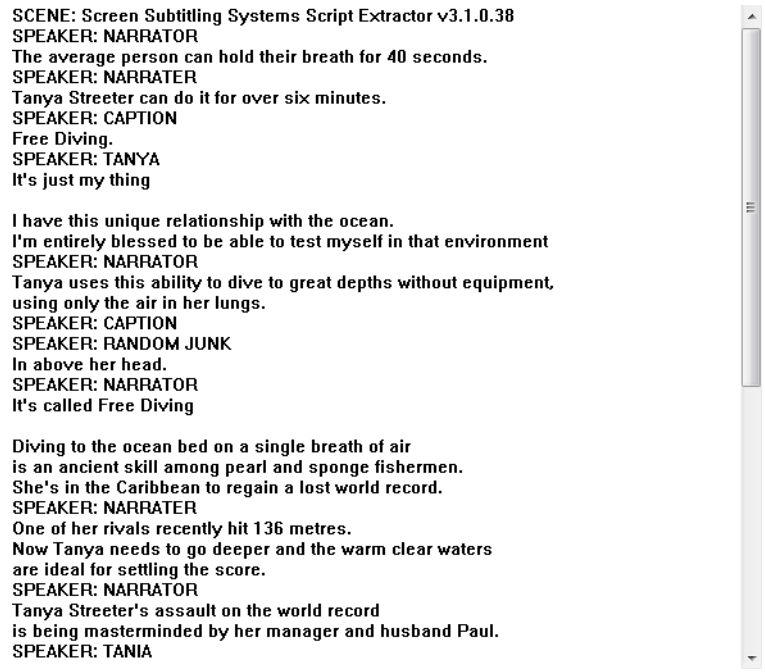
This displays the Extract Dialogue Options window.



3 Using the information below, ensure the appropriate boxes are checked to extract the required content then click  .

Option	Function	
Include Speaker ID	Speaker IDs will be included with the dialogue that you extract.	
Include Timecodes	Timecodes will be included with the extracted dialogue. The timecodes are extracted exactly as they appear in the script, unless you check one or both of the adjacent boxes.	
	Tidy	When checked, ensures that within the extracted text, timecodes all appear in the standard format (hh:mm:ss:ff) which may be a requirement for other applications used within the workflow.
	Bracket	When checked, ensures that within the extracted text, timecodes are enclosed in brackets. This may be a requirement for other applications in the workflow.
Include Scene Boundaries	Scene boundaries will be included with the dialogue that you extract.	

Script Extractor then classifies all the elements within the script, and displays in the List pane the content that will be extracted.



In the above example, only the dialogue and speaker IDs were extracted.

You are now ready to ***review and fine-tune*** (page 18) the results of the extraction.

Fine-tuning the extraction

The accuracy of the results from the first-pass extraction depends on the consistency in layout of the script and the training examples you have used. With the minimal training that is typical for a first pass, you may find some unwanted information is being included in the results, or perhaps that some wanted information is missing.

If so, you can fine-tune the extraction by providing further training examples and running the extraction again. If the results from retraining and rerunning the extraction are still not accurate, you can make final adjustments manually.

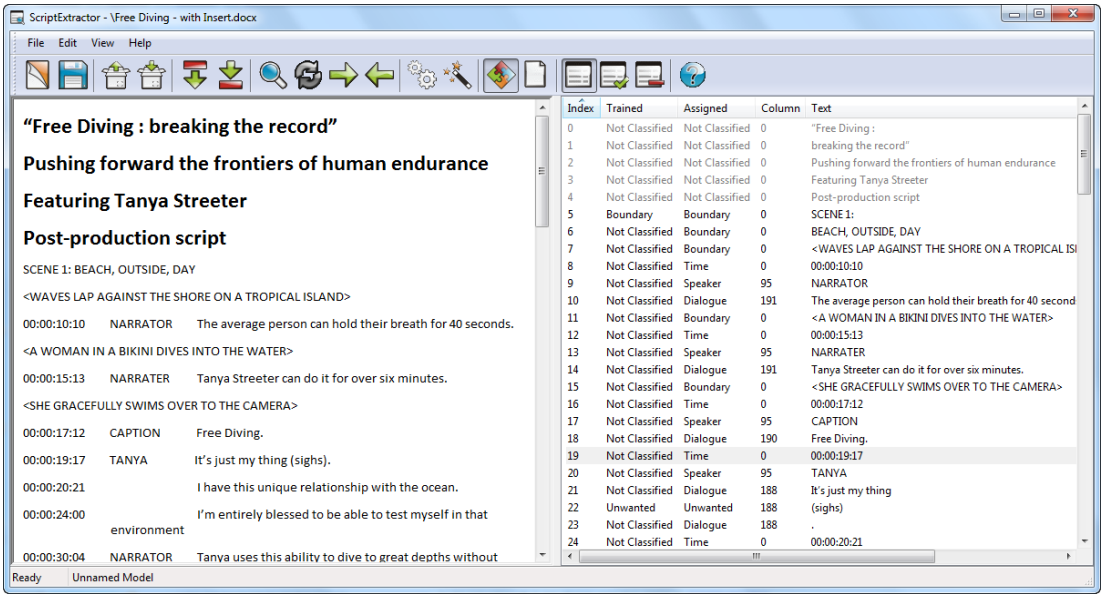
Below are guidelines on fine-tuning the extraction. For more detailed information see: ***Assessing extracted content*** (page 21) and ***Techniques for refining results*** (page 27).

To fine-tune the extraction

1 If you wish to review the classifications assigned by Script Extractor, do one of the following to display the list of items and their classification.

- Click in the toolbar.
- From the main menu, select **View>Classifications**.

- Press **F9**.



2 With the classifications displayed in the List pane, reclassify two typical items of any items that have been incorrectly classified. If you find unwanted items with a consistent style in the results, label them as 'Unwanted'.

   Notice that as you change a classification, the classification in both the 'Trained' and 'Assigned' columns becomes green.

3 Redo the ***extraction*** (page 16).

   The items you have reclassified manually are now shown in red.

When you have extracted the required items from the script, you can ***save the results*** (page 19).

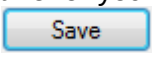
Saving the results

To save the results of the extraction

1 Do one of the following to save the extracted script items.

- Click in the toolbar.
- From the main menu, select **File>Save Dialogue**.
- Press **Ctrl+S**.

The Save As dialog is displayed.

2 Enter a name for your file and select the folder where you want it to be saved, then click 

.

Your file is by default saved as a text file (.txt), and if necessary, you can make any minor amendments to it that may be required.

Assessing extracted content

This section describes the tools that can help you to review the results of the extraction so you can assess how much further manual intervention is needed to extract the dialogue and other information you require. Some of the tools can be used in combination, so, for example, you could first ***filter*** (page 24) the classified list and then ***sort the result by column*** (page 23), or ***find*** (page 25) a word in the script then ***synchronise*** (page 21) the List pane to show the same item.

The section on ***Techniques for refining results*** (page 27) gives further information on correcting any errors you find, and general guidance on the approach to take.

Synchronising script and list items

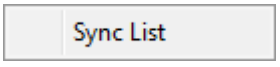
Synchronising the Script and List panes means using an item highlighted in one pane to locate the same item in the other pane. This helps you to:

- review the context of items
- assess whether items have been accurately classified and will therefore be included in the information you wish to extract.

Note: You can only synchronise the script and list items when the List pane is displaying the classifications.

To synchronise the list with the script

1 In the Script pane, use any of the following:
   - the vertical scroll bar
   - **PageUp** and **PageDown** keys
   - keyboard up and down arrow keys

   to find the item you wish to synchronise to, then double-click it to select it.
2 Do one of the following to synchronise the display in the List pane with the display in the Script pane.
   - Click in the toolbar.
   - Right-click on the selected item, then click on the 

pop-up that is displayed.
   - Press **Ctrl+D**.
   - From the main menu, select **Edit>Sync To Document**.

   You will now see the same item highlighted in both panes.

To synchronise the script with the list

1 In the List pane, locate the item you wish to synchronise to, then click to select it.

2 Do one of the following to synchronise the display in the Script pane with the display in the List pane.

- Click in the toolbar.
- Press **Ctrl+L**.
- From the main menu, select **Edit>Sync To List**.

You will now see the same item highlighted in both panes.

Showing or hiding columns

Whenever the List pane shows the classifications, the information associated with each element in the script is displayed in various columns. Four columns are always shown and their content is as follows:

- **Index** - an ascending number that identifies the order of the elements in the script.
- **Trained** - displays either 'Not Classified' or the classification that you have given to that element.
- **Assigned** - displays either 'Not Classified' or the classification assigned by Script Extractor.
- **Text** - displays the content of the element.

Other columns may be displayed or hidden as you wish. These are:

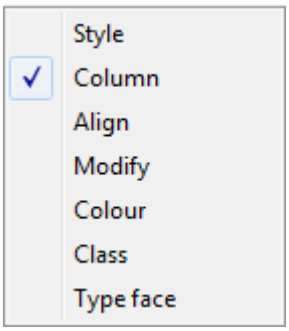
- **Style** - shows the style of the element, for example, bold, italics, underlined.
- **Column** - shows a column number derived from the first character of the item.
- **Align** - shows the alignment of the item, for example, left, right, centred.
- **Modify** - shows how text is presented in the original script.
- **Colour** - shows the colour of the item as it appears in the script.
- **Class** - displays the classifications assigned to different content fields by the script production template, if available.
- **Typeface** - shows the typeface for the item, for example, arial, verdana, times roman.

You can use any combination of available columns to ***sort*** (page 23) the items in the List pane.

To show or hide a column

1 Do one of the following to display the column settings.

- Go to the **View** menu and position the mouse over the Column option.
- Position the mouse cursor anywhere within the column headings bar and right click.



When an option is checked, that column and its associated information will be displayed in the List pane. Clicking on an option toggles between displaying and hiding that column.

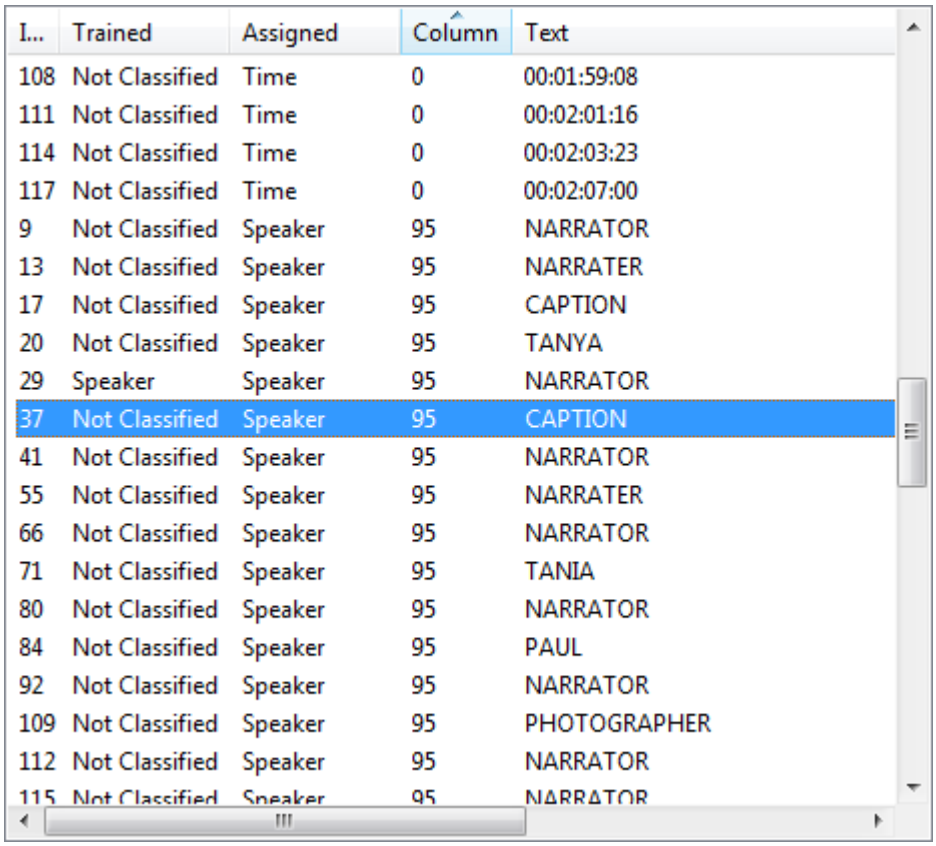
2 Within the drop-down list of options, click on the type of column you wish to hide/display.

3 Repeat the above steps to hide or display each of the optional columns as you wish.

Sorting by column

You can use the information in any of the List pane columns to group items together so that you can quickly identify any errors.

In the script used for the example below, all of the speaker IDs start in column 95 and all of the timecodes start in column 0. Sorting by column groups the elements in the script by column, and you can quickly identify that item 37 is a caption, rather than a speaker ID.

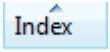


I...	Trained	Assigned	Column	Text
108	Not Classified	Time	0	00:01:59:08
111	Not Classified	Time	0	00:02:01:16
114	Not Classified	Time	0	00:02:03:23
117	Not Classified	Time	0	00:02:07:00
9	Not Classified	Speaker	95	NARRATOR
13	Not Classified	Speaker	95	NARRATER
17	Not Classified	Speaker	95	CAPTION
20	Not Classified	Speaker	95	TANYA
29	Speaker	Speaker	95	NARRATOR
37	Not Classified	Speaker	95	CAPTION
41	Not Classified	Speaker	95	NARRATOR
55	Not Classified	Speaker	95	NARRATER
66	Not Classified	Speaker	95	NARRATOR
71	Not Classified	Speaker	95	TANIA
80	Not Classified	Speaker	95	NARRATOR
84	Not Classified	Speaker	95	PAUL
92	Not Classified	Speaker	95	NARRATOR
109	Not Classified	Speaker	95	PHOTOGRAPHER
112	Not Classified	Speaker	95	NARRATOR
115	Not Classified	Speaker	95	NARRATOR

To sort by column

1 Ensure that you have ***displayed any optional columns*** (page 22) that may contain information you wish to sort on.

2 In the List pane, click on the heading of the column you wish to sort by. For example, to sort by the classification assigned by Script Extractor, click on 'Assigned'. The column heading changes to show that a sort is active:



Only one sort may be active at any time.

3 If you wish to return to the order in which the items appear in the script, simply sort by 'Index'.

Filtering results

Filtering allows you to view just the items that will be included in the output, or just the ones to be discarded. This can help you quickly decide if anything has been classified wrongly.

To filter the results

1 Ensure the classifications are displayed in the List pane.

When the classifications are shown, the filter commands become available. By default, all classified items are displayed.

2 If you wish to display only those items included in the extraction, do one of the following.

- Click in the toolbar.
- Press **Ctrl+I**.
- From the main menu, select **View>Classifications Filter>Included**.

The display in the List pane is updated to show only those items that will be included in the extraction.

3 If you wish to display only those items to be discarded, do one of the following.

- Click in the toolbar.
- Press **Ctrl+R**.
- From the main menu, select **View>Classifications Filter>Discarded**.

The display in the List pane is updated to show only those items that will be excluded from the extraction.

4 If you wish to display all classified items, do one of the following.

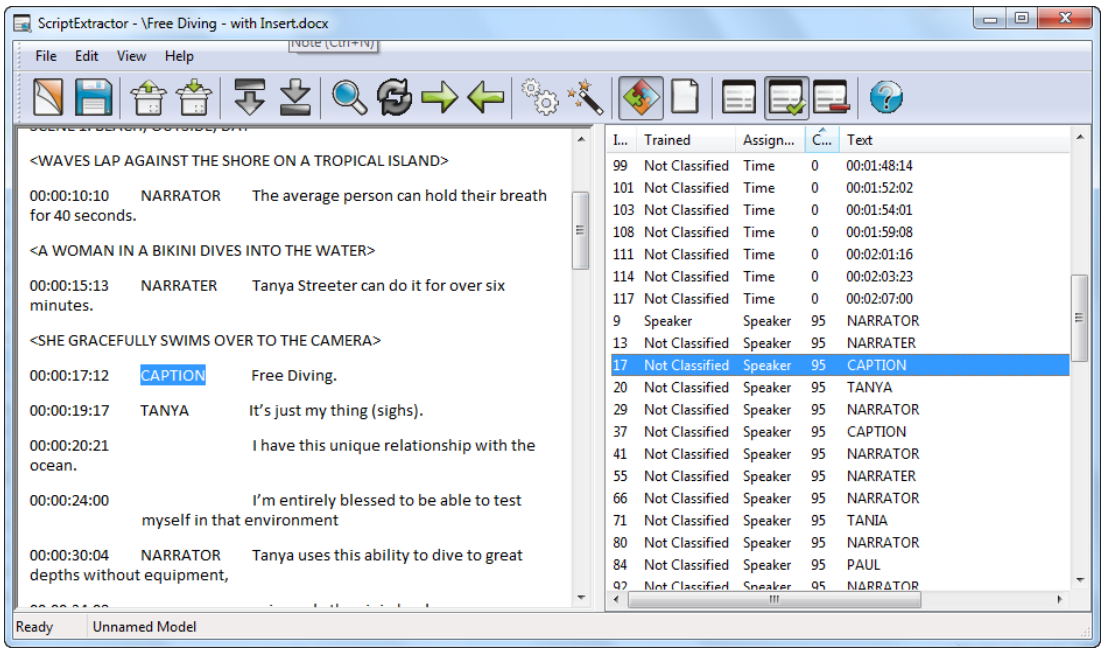
- Click in the toolbar.
- Press **Ctrl+A**.
- From the main menu, select **View>Classifications Filter>All**.

Finding a specific word or phrase

The find function allows you to locate a word or phrase within the script. This can be useful if you would like to investigate any particular item in further detail.

There is a separate ***find and replace*** (page 30) function for finding a string in the extracted content.

In the following example CAPTION has been classified as 'Speaker'.



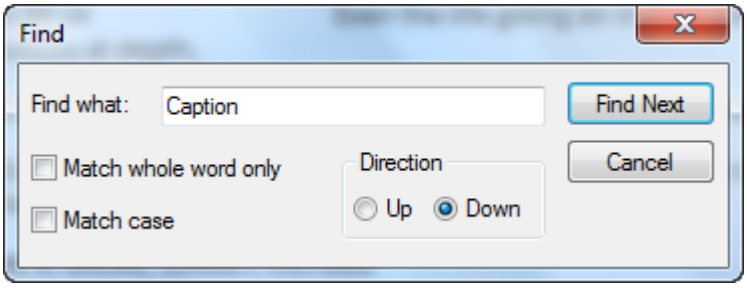
You could ***synchronise*** (page 21) the panes to display this instance of CAPTION in the script, but this would not show any other instances. In this situation it may be better to use Find.

To find one or more instances of a specific word or phrase

1 Do one of the following.

- Click in the toolbar.
- Press **Ctrl+F**.
- From the main menu, select **Edit>Find Text...**.

This displays the **Find** dialog.



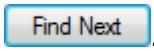
2 Enter the string you wish to search for in the **Find what:** box.

3 If you wish to match the character string exactly as you have entered it, ensure the 'Match whole word only' box is checked.

When checked, a search string of 'CAP' would not find 'CAPTION' in the script.

4 If you wish to make the search case-sensitive, ensure the 'Match case' box is checked.

When checked, a search string of 'Caption' would not find 'CAPTION' in the script.

5 Select whether you are searching up (backwards through the script) or down (forwards through the script), then click 

.

The next instance of your search string is shown highlighted in the Script pane.

Techniques for refining results

Once you have reviewed the results of the extraction, you may wish to refine the results. This involves adjusting the classification(s) assigned by Script Extractor. However, there are various ways you can do this.

Resetting the classification

If after reviewing the initial classification you identify some items which were not properly classified, it may be because you had not provided an example. For example, you may have forgotten to include an example of a boundary and all the boundaries within the script have therefore been classified as dialogue.

Where examples are missing, you can simply ***retrain*** (page 18) Script Extractor - in the above case to add examples of boundaries - and run the extraction again. This time when you review the results these items should be classified correctly.

If some classifications remain incorrect, it may be that the style or presentation of the item is not distinct enough that Script Extractor can be trained to identify it consistently. In this situation you can ***override the classification manually*** (page 27).

Manually overriding the classification

Manually overriding the classification(s) assigned by Script Extractor is a quick way of fine-tuning the results when the extraction is not totally accurate. A manual override changes the classification of specific items without interfering with the training of any other items.

Note: If you go on to ***create a model*** (page 33) from the training, any manual changes to the classifications will not be included in the model. This is because manual overrides are likely to be specific to a script and therefore do not contribute to the training.

When manually overriding the assigned classification, you can change a single item, or group similar items together and change their classification in one go. If you make a mistake when overriding the classification, you can ***undo*** (page 29) any manual changes you have made.

To change the classification of a single item

1. Ensure the classifications are displayed in the List pane, then locate the item whose classification you wish to override.
2. In the List pane, click on the item to select it.
3. When the item is selected, hold down the **Shift** key and press the appropriate key from the following.
   - S - Speaker ID

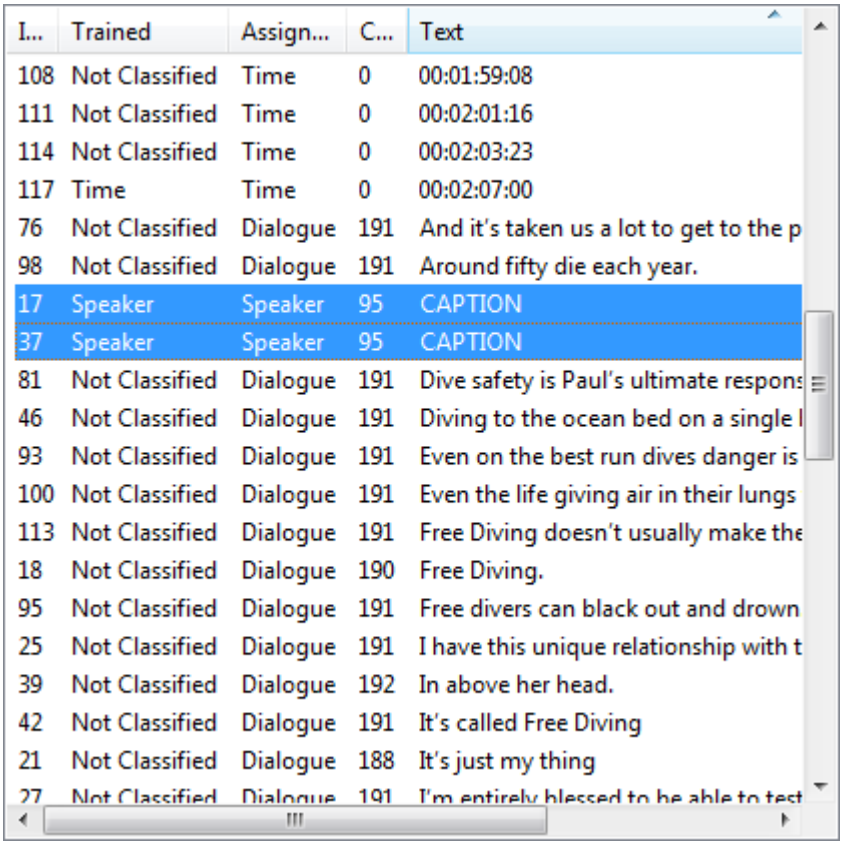
- D - Dialogue
- T - Timecode
- B - Boundary
- P - Production
- U - Unwanted

The item you have changed now shows the new classification in both the 'Trained' and 'Assigned' columns in mauve.

To change the classification of multiple items

1 If required, ***sort*** (page 23) the items to group together the ones you wish to change.

In the example below, the items have been sorted by Text so that the misclassified 'CAPTION's appear together.



I...	Trained	Assign...	C...	Text
108	Not Classified	Time	0	00:01:59:08
111	Not Classified	Time	0	00:02:01:16
114	Not Classified	Time	0	00:02:03:23
117	Time	Time	0	00:02:07:00
76	Not Classified	Dialogue	191	And it's taken us a lot to get to the p
98	Not Classified	Dialogue	191	Around fifty die each year.
17	Speaker	Speaker	95	CAPTION
37	Speaker	Speaker	95	CAPTION
81	Not Classified	Dialogue	191	Dive safety is Paul's ultimate respons
46	Not Classified	Dialogue	191	Diving to the ocean bed on a single l
93	Not Classified	Dialogue	191	Even on the best run dives danger is
100	Not Classified	Dialogue	191	Even the life giving air in their lungs
113	Not Classified	Dialogue	191	Free Diving doesn't usually make the
18	Not Classified	Dialogue	190	Free Diving.
95	Not Classified	Dialogue	191	Free divers can black out and drown
25	Not Classified	Dialogue	191	I have this unique relationship with t
39	Not Classified	Dialogue	192	In above her head.
42	Not Classified	Dialogue	191	It's called Free Diving
21	Not Classified	Dialogue	188	It's just my thing
27	Not Classified	Dialogue	191	I'm entirely blessed to be able to test

2 To select consecutive lines within the list, do the following.

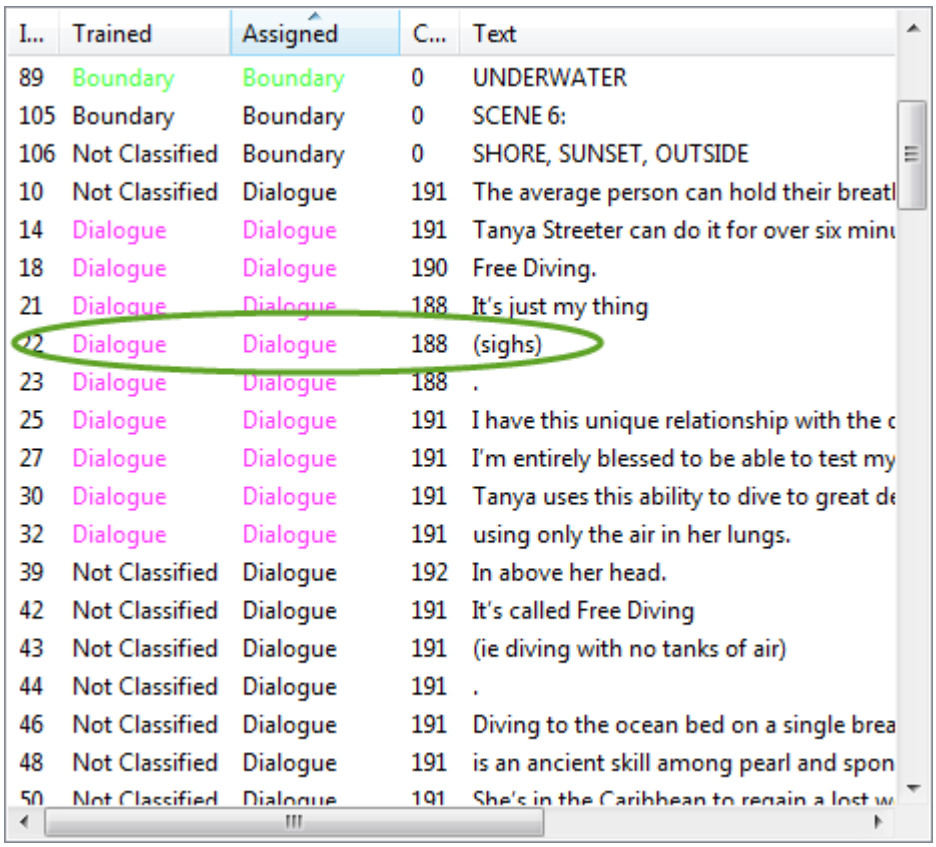
- Select the first (or last) item by clicking on it.
- Hold down the **Shift** key and click on the item that completes the selected block.

All lines between the items you have clicked will be selected.

3 Hold down the **Shift** key and press the appropriate classification key to manually reclassify all of the selected items.

Undoing a classification

If you have manually overridden the classification of an item, you can reset it to the previous classification by undoing your override. This can be useful if you have changed the classification for a block of items and then wish to change just one of them back, for example, as shown below.



I...	Trained	Assigned	C...	Text
89	Boundary	Boundary	0	UNDERWATER
105	Boundary	Boundary	0	SCENE 6:
106	Not Classified	Boundary	0	SHORE, SUNSET, OUTSIDE
10	Not Classified	Dialogue	191	The average person can hold their breatl
14	Dialogue	Dialogue	191	Tanya Streeter can do it for over six minu
18	Dialogue	Dialogue	190	Free Diving.
21	Dialogue	Dialogue	188	It's just my thing
22	Dialogue	Dialogue	188	(sighs)
23	Dialogue	Dialogue	188	.
25	Dialogue	Dialogue	191	I have this unique relationship with the c
27	Dialogue	Dialogue	191	I'm entirely blessed to be able to test my
30	Dialogue	Dialogue	191	Tanya uses this ability to dive to great de
32	Dialogue	Dialogue	191	using only the air in her lungs.
39	Not Classified	Dialogue	192	In above her head.
42	Not Classified	Dialogue	191	It's called Free Diving
43	Not Classified	Dialogue	191	(ie diving with no tanks of air)
44	Not Classified	Dialogue	191	.
46	Not Classified	Dialogue	191	Diving to the ocean bed on a single brea
48	Not Classified	Dialogue	191	is an ancient skill among pearl and spon
50	Not Classified	Dialogue	191	She's in the Caribbean to regain a lost w

The items in mauve have been overridden as a block, but the single item identified above may need to be reclassified with a further override, or returned to its previous classification.

Undoing an override is not cumulative and will reset only to the immediately previous classification.

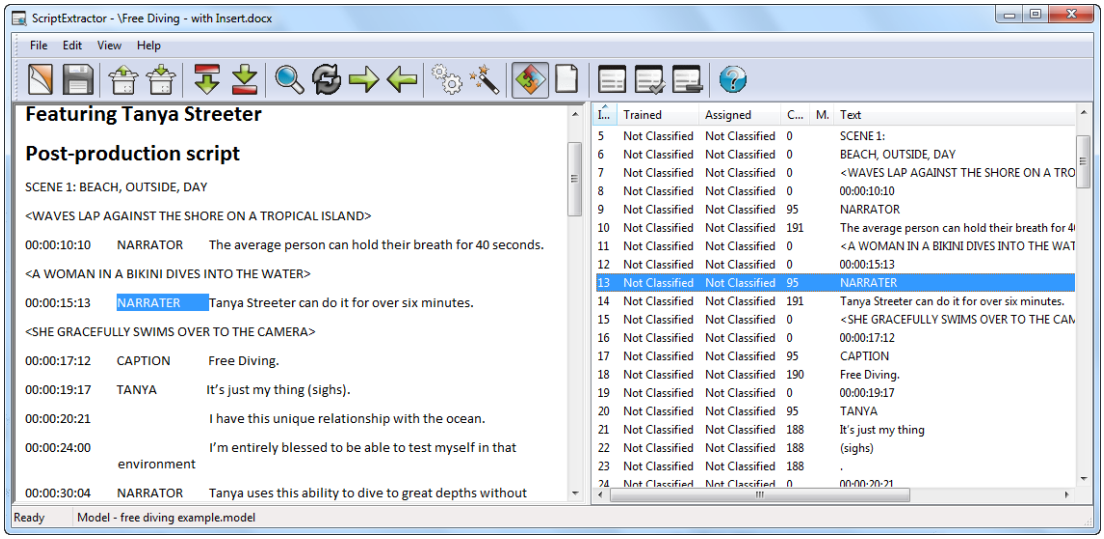
To undo an override

1. Select the individual item, or block of items to be put back to the previous classification.
2. Press the **Space** bar.

Finding and replacing text

The find and replace function allows you to locate a word or phrase within the extracted content and optionally replace it. This can be useful if you notice errors in the extracted content and would like to correct them.

In the following example NARRATOR has been misspelled as 'NARRATER'.



Find and replace allows you to find all occurrences of 'NARRATER' in the extracted content and replace them with 'NARRATOR'.

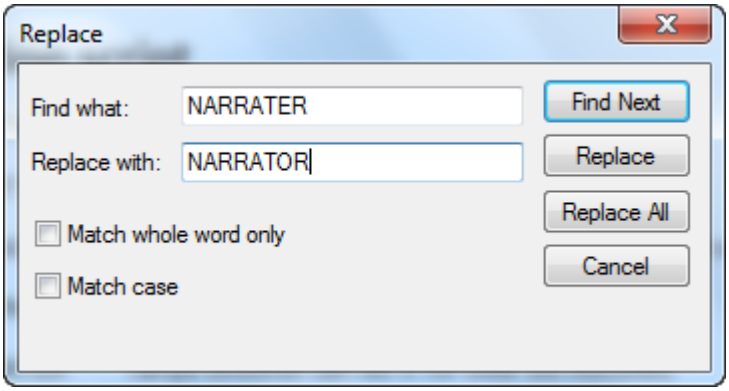
To find and optionally replace one or more instances of a specific word or phrase

1 Ensure that the List pane is set to 'Show Dialogue' to display the results of the last extraction.

2 Do one of the following.

- Click in the toolbar.
- Press **Ctrl+H**.
- From the main menu, select **Edit>Replace Text...**.

This displays the **Replace** dialog.



3 Enter the string you wish to search for in the **Find what:** box.

4 If you wish to match the character string exactly as you have entered it, ensure the 'Match whole word only' box is checked.

When checked, a search string of 'NAR' would not find 'NARRATER' in the script.

5 If you wish to make the search case-sensitive, ensure the 'Match case' box is checked.

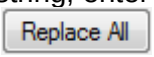
When checked, a search string of 'Narrater' would not find 'NARRATER' in the script.

6 Click

.

The next instance of your search string is shown highlighted in the extracted dialogue shown in the List pane.

7 If you wish to replace the string, enter the new string in the **Replace with:** box and click either

or 

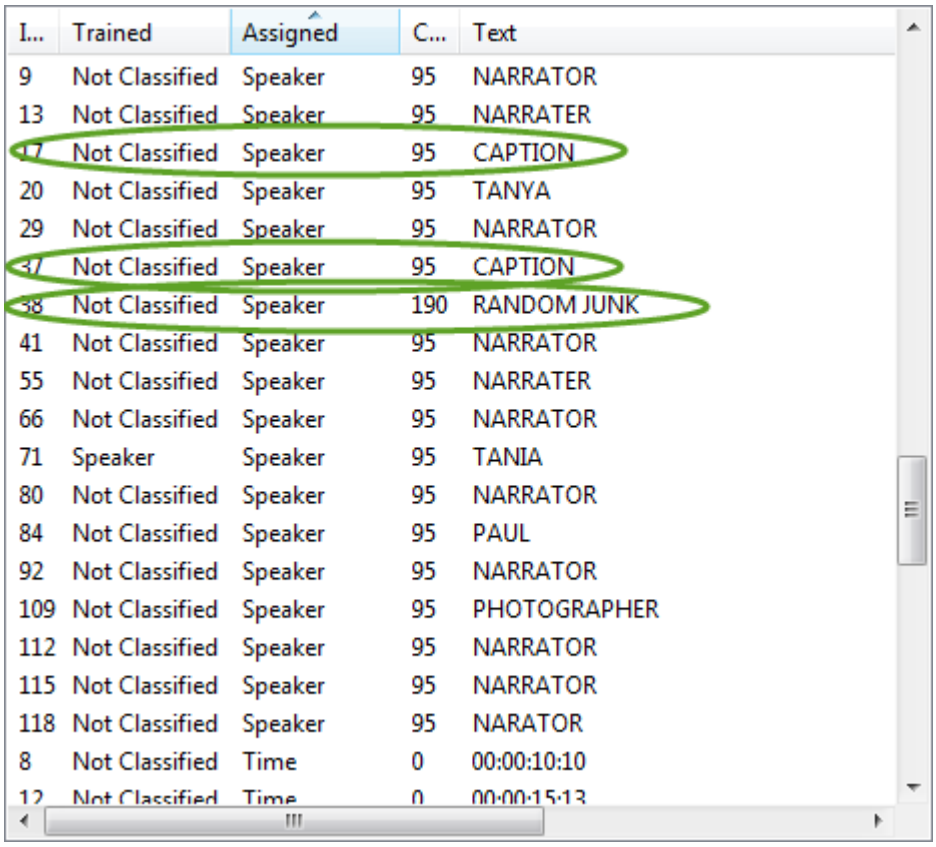
as appropriate.

Deciding the most effective approach

Generally, if the script has a consistent layout, it is always more efficient to ***train Script Extractor*** (page 15) so that items are classified automatically. However, there may be situations where different items are similar (either in their style or positioning) which makes it difficult for Script Extractor to distinguish the two different types.

Where this is the case, retraining Script Extractor may result in an even less accurate classification, and it may be better to ***reassign the classifications manually*** (page 27).

In the example below, two instances of 'CAPTION', and 'RANDOM JUNK' have been misclassified as speakers. The former start in column 95 the same as speaker IDs in the script. If you were to label the 'CAPTION' items as 'Unwanted' and retrain Script Extractor, for example, you may also find the speaker IDs being reassigned as unwanted.



I...	Trained	Assigned	C...	Text
9	Not Classified	Speaker	95	NARRATOR
13	Not Classified	Speaker	95	NARRATER
17	Not Classified	Speaker	95	CAPTION
20	Not Classified	Speaker	95	TANYA
29	Not Classified	Speaker	95	NARRATOR
37	Not Classified	Speaker	95	CAPTION
38	Not Classified	Speaker	190	RANDOM JUNK
41	Not Classified	Speaker	95	NARRATOR
55	Not Classified	Speaker	95	NARRATER
66	Not Classified	Speaker	95	NARRATOR
71	Speaker	Speaker	95	TANIA
80	Not Classified	Speaker	95	NARRATOR
84	Not Classified	Speaker	95	PAUL
92	Not Classified	Speaker	95	NARRATOR
109	Not Classified	Speaker	95	PHOTOGRAPHER
112	Not Classified	Speaker	95	NARRATOR
115	Not Classified	Speaker	95	NARRATOR
118	Not Classified	Speaker	95	NARATOR
8	Not Classified	Time	0	00:00:10:10
12	Not Classified	Time	0	00:00:15:13

In this instance, because there are so few incorrect classifications it is better to reassign the classifications manually so as not to interfere with the other, correct classifications.

Using models

Models are a way to store the Script Extractor training (the training examples you have used), so that the next time you process a script with the same layout you can reuse the training examples. This saves time because once you have loaded both the script and the model, you can skip the training and go straight to the extraction.

Models are most effective when the scripts to be processed have exactly the same layout - you can use any of your saved models on any script, but if the training in the model is not relevant to your script layout, the results will not be accurate.

The status bar indicates the name of the model currently being used, or 'No model' when no model is applied.

Creating a new model

You can create a model as an optional step any time after you have completed the Script Extractor training. You do not need to have ***saved the results*** (page 19) of the extraction before you create a model.

To create a new model

1. Load the script, and ensure that 'No model' is displayed in the status bar.
   If an existing model is shown you will need to ***clear*** (page 35) it.
2. Ensure that you have completed the Script Extractor ***training*** (page 15) relevant to the script you are working on.
3. Do one of the following.
   - Click in the toolbar.
   - From the main menu, select **File>Save Model**.
   - Press **Ctrl+P**.

   The Save As dialog is displayed.
4. Enter a name for your file and select the folder where you want it to be saved, then click

.
   Your file is by default saved as a model file (.model) and you can ***use*** (page 34) it the next time you load a script of the same type. The status bar is updated to show the name of the saved model.

Using a saved model

It does not matter whether you load the script first then load the model or do it the other way round. Similarly, if you wish to ***top and tail*** (page 14) the script, you can do this either before or after you have loaded the model (but before you extract the useful script elements).

Once the model is loaded, it will be used when you extract script items. Following the extraction, you may still need to make manual adjustments to the extracted content, such as ***manual overrides*** (page 27), to obtain only the items you require.

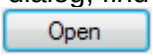
To use a pre-defined model to extract information from a new script

1 ***Load the script*** (page 13) as normal.

2 Do one of the following.

- Click in the toolbar.
- From the main menu, select **File>Load Model**.
- Press **Ctrl+G**.

This launches the **Open** dialog.

3 Within the **Open** dialog, find and select the file containing the model you wish to load, then click 

.

The name of your selected model is loaded and shown in the status bar.

4 ***Extract*** (page 16) the script items as normal, fine-tuning the results as required.

Updating an existing model

There may be situations where you need to modify a model that you have previously saved, for example, if

- the layout of the modelled script changes to become more consistent. This may mean that script elements that you previously needed to manually override can now be recognised automatically
- you receive a new script with a layout similar but not identical to one you already have a model for.

In such cases you can use an existing model, update it and save the changed model either with the same name, or with a different name if you wish to keep both models.

To modify an existing model

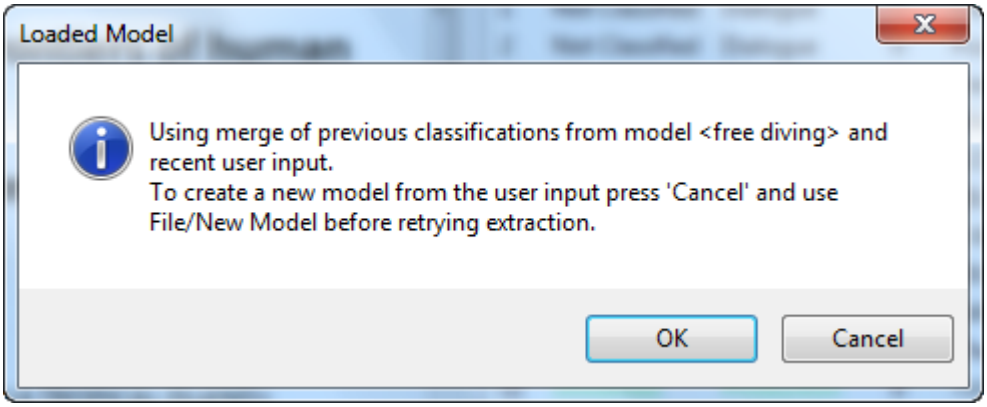
1 Load the model you wish to modify and the script on which the new model is to be based.

2 If you wish to review the results of the existing model on the loaded script, extract the content as usual.

3 Train Script Extractor on the script elements you wish to be included in the new model.

4 Run an extraction to ensure that the new model gives the results you want.

Before Script Extractor processes the script you will see the following dialog.



5 If you wish to use and save the updated model, click OK, then when the extraction is complete, save the model by doing one of the following.

- Click in the toolbar.
- From the main menu, select **File>Save Model**.
- Press **Ctrl+P**.

The Save As dialog is displayed. Using the same model name will overwrite the existing model, or enter a different name to create a new model.

6 If you wish to create a completely new model based only on the training for this script, ***clear the model*** (page 35), then ***create a new model*** (page 33).

Clearing a model

If during the current session you have used a model when working on a script, that model will remain in force until you either load a new model or specifically clear it. This means that if you open a new script, any model currently loaded will be applied to that script when you extract the dialogue.

Note: The status bar shows the model, if any, that is currently loaded.

If the new script has the same layout and style as the previous script you can continue to use the same model. However, if the model is not suitable for the new script, you will either have to use another model more suited to that script, or to clear the model. Clearing the model will then allow you to train Script Extractor in the normal way. *If you forget to clear the model, you will get poor results.*

To clear a model

Do one of the following.

- From the main menu, select **File>New Model**.
- Press **Ctrl+N**.

The status bar is updated to display 'No Model'.

Working in batch mode

If you have a number of scripts to work on all of which use the same model, you can process them in batch mode. You do this by specifying

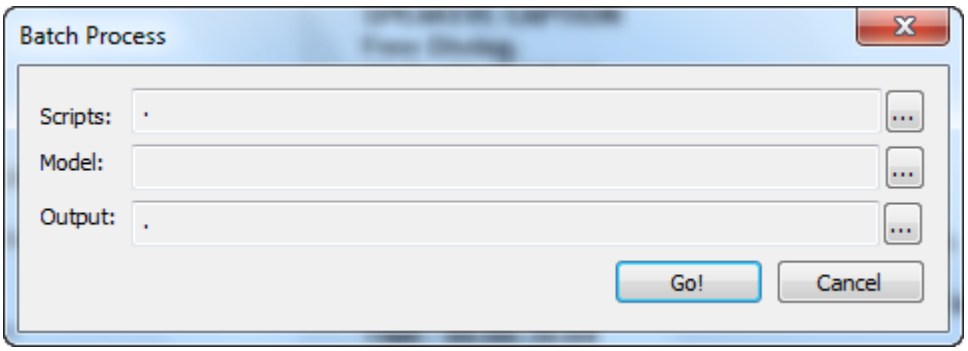
- the folder where the input script files are stored
- the name of the model to be used, and
- the folder where the output .txt files are to be saved.

Note: Batch processing does not allow you to ***top and tail*** (page 14) the script.

To process a batch of script files

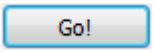
1 Do one of the following.
   - From the main menu, select **File>Batch Process**.
   - Press **F5**.

   This launches the **Batch Process** dialog.



2 Enter the file path for the input scripts, model file to be used and output folder in the appropriate boxes.

   Clicking displays the **Select Folder** dialog where you can select the required location(s).

3 Click  to start processing.

Configuring Script Extractor

You can configure Script Extractor as described in the topics that follow. The exact configuration options depend on the way Script Extractor has been set up, so not all options may be available on your system.

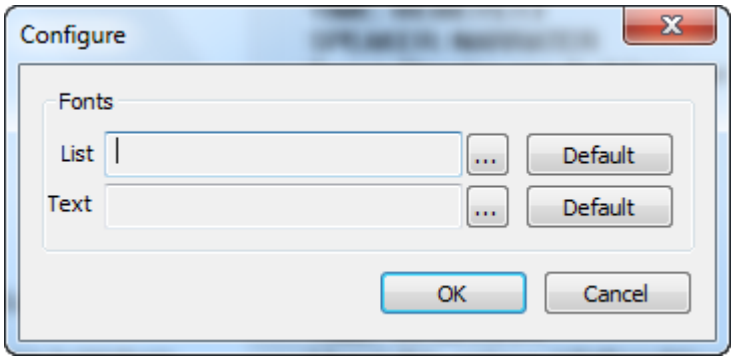
To change the way Script Extractor is configured

1 Do one of the following.
   - From the main menu, select **File>Configure**.
   - Press **Ctrl+K**.

   The Configure dialog is displayed.
2 Using the information in the topic on the ***configuration options*** (page 37), make the required changes to the configuration.
3 If you wish to return to the default setting for any of the configuration options, click Default next to that option.
4 When you have finished configuring Script Extractor, click OK to save the configuration and close the Configure dialog.

The configuration options

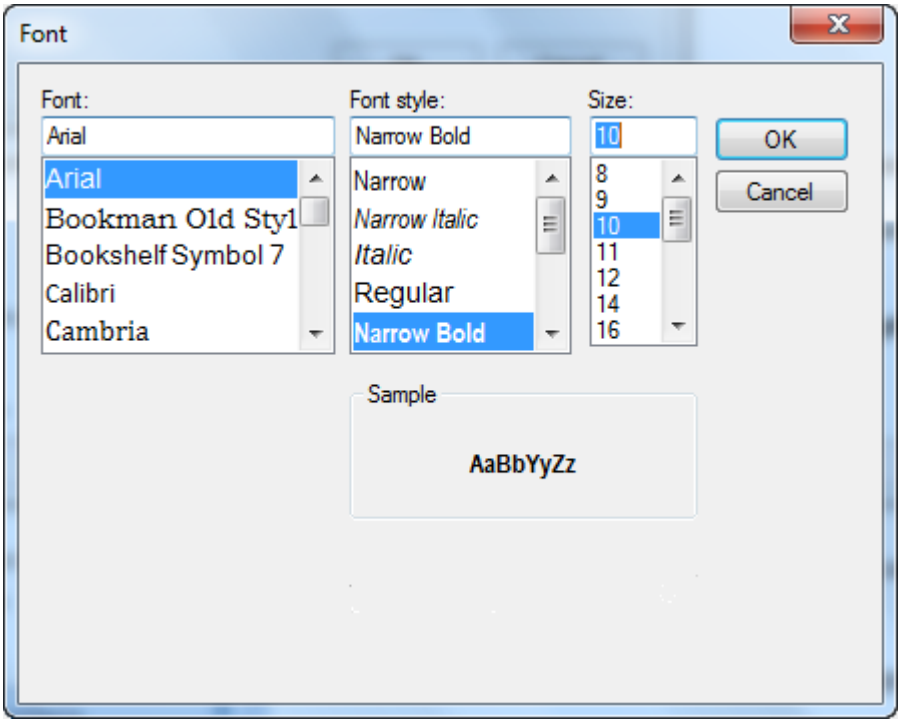
The configuration options allow you to change the way Script Extractor presents information in the ***List pane*** (page 10).



Parameter	Setting
Fonts	
List	Sets the font style to be used whenever the List pane displays dialogue.
Text	Sets the font style to be used whenever the List pane displays classifications.

To select the font

1 Click ... to display the Font dialog where you can select the required font.



2 Using the drop-down lists, select the font name, style and size you require. Your selection is previewed in the Sample box.

3 When you have made your selection, click OK to return to the Configure dialog.

4 If you wish to return to the default setting, click the appropriate Default button in the Configure dialog.

Troubleshooting and tips

If you are not getting the results you were expecting from Script Extractor, the following guidance may help you to understand why. If you continue to experience problems using the product, contact the Screen Systems customer support team for assistance.

Depending on your installation, you may find additional help is available from the local help page button, in the toolbar.

Troubleshooting

Symptom	Remedy
Script Extractor doesn't work at all	Ensure that you have a licence for the product installed and that Microsoft Word is also installed and licensed on your computer.
With a script loaded there is no information in the List pane	The most likely explanation is that the List pane is set to display dialogue and you have not yet extracted any script items. One of the following should resolve this. Click in the toolbar and you should now see the script elements listed in the List pane. Alternatively, if you are ready to extract the required items, click in the toolbar and the extracted text should appear in the List pane.
The classifications are not displayed	The List pane is probably set to display the dialogue. Click in the toolbar and you should now see the script elements and their classifications in the List pane.
The List pane shows classifications, but some of the columns are missing	Ensure that you have set your display so that all the information you want to see is displayed. See: ***Showing or hiding columns*** (page 22) for more information.

The information in the List pane is all greyed out	If you have attempted to ***top and tail the script*** (page 14), you have probably forgotten to synchronise the List pane with the script. In the Script pane, simply click on any word on the last line of the script proper, click in the toolbar, then click . This will set the end point of script and grey out only the items below your selected end point.
The timecodes and/or scene boundaries are not showing in the dialogue	This could be due to either of the following. ▪ The List pane is currently set to display dialogue and you did not select these items to be included on extraction. Redo the extraction, this time ensuring that all the items you wish to extract from the script are checked in the **Extract Dialogue Options** dialog. See: ***Extracting the required content*** (page 16) for more information. ▪ The List pane is currently set to display script elements and classifications, but the filters are set so that the items you are expecting to see are not displayed. See: ***Filtering results*** (page 24) for more information.
The results of the extraction are not as expected	This could occur for either of the following reasons. ▪ The training examples are not typical or are otherwise not representative of the script elements you wish to train Script Extractor to recognise. Try providing one or two more examples to ***retrain*** (page 15) Script Extractor and run the extraction again. ▪ A model may already be loaded when you load a new script. If the model is not appropriate for the script you may get inaccurate results. ***Clear the model*** (page 35) if there is one loaded, then either ***train*** (page 15) Script Extractor or ***load a model*** (page 34) more suited to the script.

Tips

The following advice may help you understand more quickly how to use the product.

Choosing training examples - Script Extractor works best when the examples you give are 'representative'. So for example, when providing examples of speaker ID, use a different speaker name for each example. If you are giving examples of dialogue, avoid one-word dialogue items and unusual items like a SHOUTED word. Choose typical examples.

If there are some consistent unwanted items ('CUT TO:' is a good example), train two or three of them as Unwanted so that Script Extractor can distinguish these from wanted script items.

How do you decide when retraining is necessary? It is largely a matter of judgment when to retrain Script Extractor and when to change classifications manually. See: ***Assessing extracted content*** (page 21) and ***Techniques for refining results*** (page 27) for more guidance.

Don't spend too much time training Script Extractor - it is probably more efficient overall to use Script Extractor to get you most of the way towards extracting the content you are interested in and then making final adjustments manually. Overdoing the training, especially if the script layout is inconsistent or the examples are not distinctive enough, can be counter-productive.

What's the best way to review the results? ***Sorting*** (page 23) can be helpful to assess (for example) whether all the speaker labels fall into the same column in the script. Once you have sorted the script items (using the most appropriate column on which to sort) you can then scan down the results to see if there are any errors or differences.

Use the filters (page 24) to restrict the displayed list of classified items to show only those items that will be included in the extracted set, or just the items that will be discarded. By scanning these lists, you can quickly decide if anything has been classified wrongly.

The Script Extractor toolbar and menus

Depending on your installation, some of the toolbar and menu options described below may not be available. Similarly, the available commands may be limited (greyed out) if they are temporarily not appropriate for where you are in the workflow.

Toolbar icon	Menu option	Keyboard control	Description
	File>Open	**Ctrl+O**	Open script file. See: ***Loading a script*** (page 13)
	File>Save Dialogue	**Ctrl+S**	Save the extracted script items. See: ***Saving the results*** (page 19)
	File>Load Model	**Ctrl+G**	Load model. See: ***Using a saved model*** (page 34)
	File>Save Model	**Ctrl+P**	Save the Script Extractor training examples as a model. See: ***Creating a new model*** (page 33)
	Edit>Mark Begin	**Ctrl+B**	Mark the start of useful script. See: ***Topping and tailing the script*** (page 14)
	Edit>Mark End	**Ctrl+E**	Mark the end of useful script. See: ***Topping and tailing the script*** (page 14)
	Edit>Find Text...	**Ctrl+F**	Search for a text string in the script. See: ***Finding a specific word or phrase*** (page 25)
	Edit>Replace Text...	**Ctrl+H**	Search for a string within the extracted dialogue and optionally replace it with a new string
	Edit>Sync To Document	**Ctrl+D**	Synchronise the display in the List pane with the display in the Script pane. See: ***Synchronising script and list items*** (page 21)
	Edit>Sync To List	**Ctrl+L**	Synchronise the display in the Script pane with the display in the List pane. See: ***Synchronising script and list items*** (page 21)
	File>Process	**F6**	Reprocess the currently loaded file. The effect is to start over with the currently loaded script and model (if any).
	File>Extract Dialogue	**F7**	Extract the dialogue and other useful items from the script. See: ***Extracting the required content*** (page 16)

	View>Classifications	**F9**	Display the script items with their classifications in the List pane. When this control is active the filter settings (see below) control what data is shown
	View>Dialogue	**F10**	Display the extracted items (dialogue and other required content) in the List pane
	View>Classifications Filter>All	**Ctrl+A**	Display all script items with their classifications in the List pane
	View>Classifications Filter>Included	**Ctrl+I**	Display only those script items that will be extracted in the List pane along with their classifications
	View>Classifications Filter>Discarded	**Ctrl+R**	Display only those script items that will be omitted from the extraction in the List pane along with their classifications
			Access localised help, if available
	File>Recent Files		Display a list of all the script files you have loaded during the current session. If you wish to reload one of these files, simply click on it within the list
	File>Batch Process	**F5**	Process a number of similar scripts together as a batch. See: ***Working in batch mode*** (page 36)
	File>Metadata	**Ctrl+M**	Not available in all versions of Script Extractor.
	File>New Model	**Ctrl+N**	Clear the current model. See: ***Clearing a model*** (page 35)
	File>Lookup Model	**Ctrl+U**	Not available in all versions of Script Extractor.
	File>Save Training Data	**Ctrl+T**	Advanced function. This is not generally used, but the Screen Systems support team may ask you to use this to provide additional technical information.
	File>Configure	**Ctrl+K**	Configure the way Script Extractor is presented. See: ***Configuring Script Extractor*** (page 37)
			The following three 'Break' options force Script Extractor to take a new script element.
	Edit>Break On Bracket		When selected, an opening bracket denotes the start of a new script element.
	Edit>Break On Capitalised Sequence		When selected, a word or group of words in uppercase letters denotes the start of a new script element.
	Edit>Break Out Timecodes		When selected, a timecode denotes the start of a new script element.
	View>Toolbar		Toggle between hiding/displaying the toolbar

View>Status Bar	Toggle between hiding/displaying the status bar
View>Column	Set which optional columns are displayed/hidden when the classifications are shown in the List pane.