iManage
Making knowledge work

Training an Extractor

There are couple of ways to extract data from documents with Extract. You can choose to use one of the internally curated **Pre-Trained extractors** if that suits your use case or you can choose to **train your own extractor**. Training extractors is one of the few key aspects of a successful data extraction process. If you decide to train your own extractor, you will have few more options to choose from - **Machine Learning extractor** with either **Quick** or **Advanced Learning** locator or a **Rules-based extactor**. This quick start guide will focus on the creation and training of a **Machine Learning extractors** with a **Quick Learning locator**.

NOTE: You can view **Training Mode** only if you are assingled **Trainer** or **Superuser** privilege.

1. The Preview pane

The **left pane** is used for viewing the document and is called the **Preview pane**.

Here you will also find other features such as: **search within the document, window configuration**, **page navigators** and **document preview settings**.

2. The Training pane

The **right pane** consists of few tabs - **Review Forms**, **Tables** and **Details**.

From the **Review Forms** tab you will be able to to select **Training mode,** where you can train your extractor.

3. Document Navigator

You can use to **document navigator** at the top right of the screen to switch between different documents within that particular project.

Training an Extractor (continued)

Before you start training your extactor, it is good to understand and freely navigate all the available features which make the two panes in the **Documents training window**.
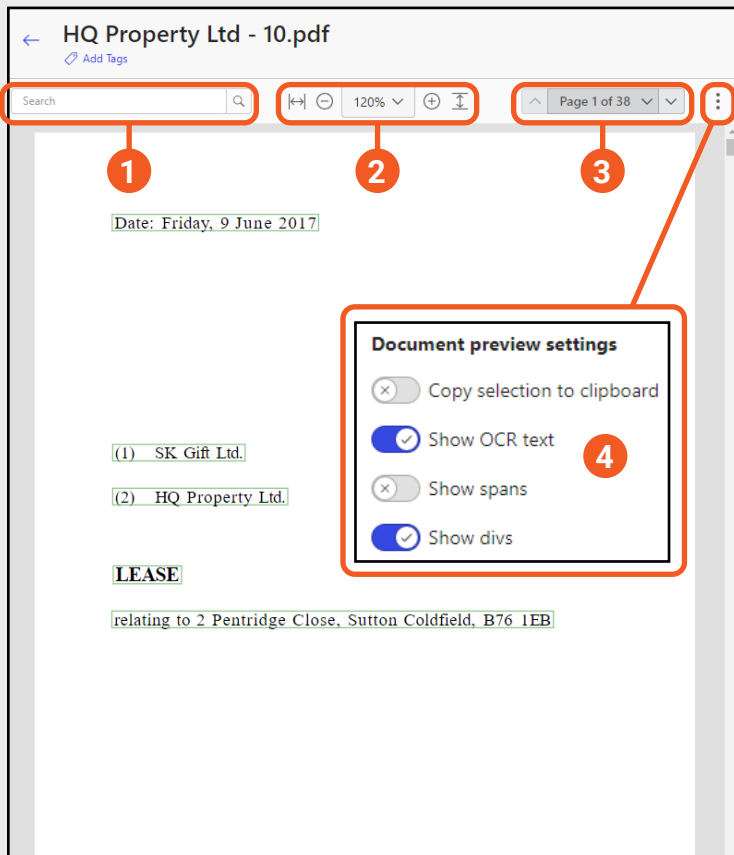
1. **Find in document**

   Search for keywords or phrases in a document using this feature. Clicking inside the search field displays the five most recent searches made by you.

2. **Fit to window & Zoom in/Zoom out**

   Use the buttons to **enlarge** the document to fit the preview window either **horizontally** or **vertically**.

   You can also **zoom in** and **zoom out** the document within the preview window or select an option from the drop-down list to **magnify** the document to the required size within the preview window.

3. **Page navigators**

   Navigate page by page or navigate to a specific page in the document.

4. **Document preview settings**

   Copy selection to clipboard: Disabled by default. If enabled, copies the content of the document outside of the iManage Extract application.

   Show OCR text: Enabled by default. If enabled, you can view the OCR version of the document, else you can view the original document.

   Show spans: Disabled by default. If enabled, each line (span) of the document is shown in the preview window.

   Show divs: Disabled by default. If enabled, each paragraph (div) of the document is shown in the preview window.

Training an Extractor (continued)

Before you start training your extactor, it is good to understand and freely navigate all the available features which make the two panes in the **Documents training window**.

1. Review Froms, Tables and Details tabs

Review Forms: The default tab where you can see choose the Review Form to work with; view, train and configure extractors; and extract data from the documents.

Tables: One of the steps during ingestion is to extract tables from documents. You can view these on the Tables tab in the left pane on the **Document Review** page.

Details: This tab on the Document Review page displays additional information about the document. You can view the following metadata values: Title, File Type, Page Count, Error Type, Error Message.

2. Select Review From

Select a **Review Form** to see its extractors and either choose to **train** or **configure** an extractor or see the extraction results of an already processed extractor.

NOTE: This drop-down shows only the already available Review Forms from that particular project.

3. Extract All button

Selecting this button will process **all** the extractors under the chosen Review Form on that **one** document.

4. Training Mode & Review Mode

Switch between **Training Mode** and **Review Mode** tabs. In **Training Mode** you will be able to **train** and **configure** extractors if you have **Trainer** or **Superuser** privileges. In **Review Mode** you will be able to **review** and **finalise** already extracted data if you have **Reviewer** or **Superuser** privileges.

NOTE: If you only have either **Trainer** or **Reviewer** priviliges, you will be able to see only the mode you have access to.

5. Extractors View

View all the extractors under a particular Review From. You can **add** or **remove** samples to any extractor, **train** the extractor, **view** already added samples, **view** extracted data points, and **configure** or **run** any single extractor.
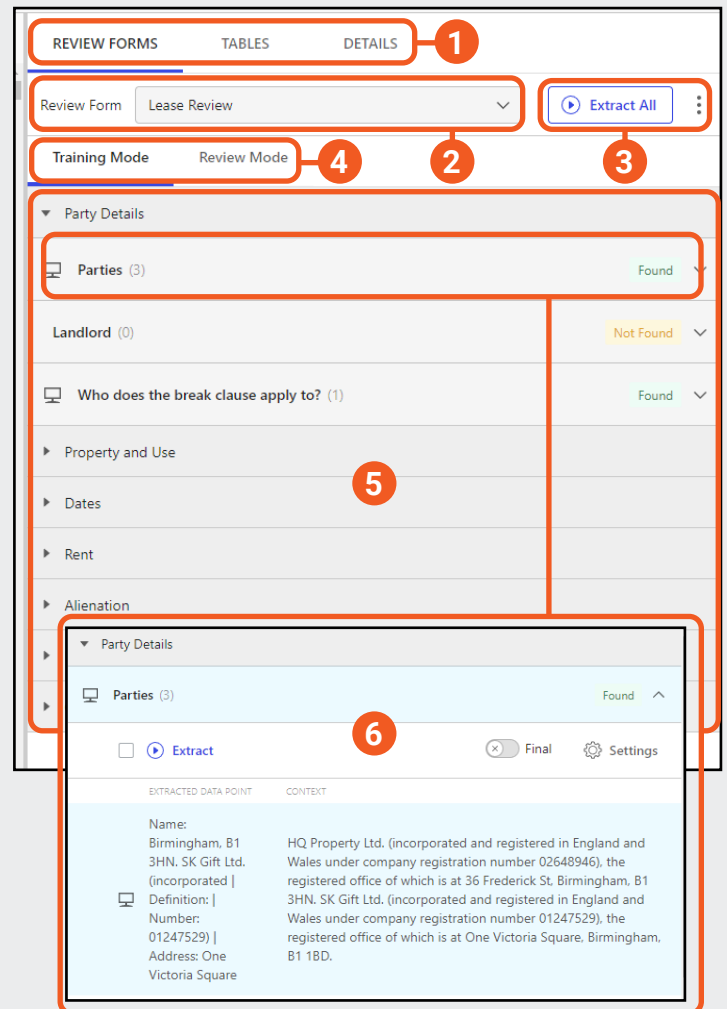
6. Single Extractor View

Selecting over an extractor will show you the details of that particular extractor. From here, you will be able to **add a sample** if the extractor has no samples, **view current samples** and **review extracted data point**, **run** that extractor over the document using the **Extract 'play'** button or **configure** the extractor by selecting the **Settings** button.

iManage
Making knowledge work

Training an Extractor (continued)

After getting familiarised with the **Documents Review** window and all the features, it is time to **train** your **Machine Learning** extractor.

1. Find in Document

Using the **Search** function, locate in the document, the **sample containing** the **data point** you would like to extract.

2. Select Sample

Drag across the required bit of text to **select** your sample. On releasing, the sample will be **automatically** added to the extractor you have chosen.

3. Extractor View before adding a Sample

If you do not have any samples or extacted data points for a particular extractor, you will see an empty field with a **Not Processed** badge on the right next to the extractor name.

4. Extractor View after adding a Sample

When you have added a sample from the document or a sample has already been added/extracted, you will see the sample itself under **Context** and the data point under **Extracted Data Point**. The badge will change to **Trained**.

5. Training Completion

After adding a sample, you can proceed to the next document and select the **Extract 'play'** button. This should automatically extract your data point. In that case, the badge will change as **Found** if a data point is found or **Not Found** if a data point has not been found.