

# Konfuzio API

## Overview

### Version information

*Version* : v2

### Contact information

*Contact Email* : [info@konfuzio.com](mailto:info@konfuzio.com)

### License information

*License* : Helm & Nagel GmbH

*Terms of service* : <https://konfuzio.com/impressum/>

### URI scheme

*Host* : app.konfuzio.com

*BasePath* : /api

*Schemes* : HTTPS

### Consumes

- `application/json`

### Produces

- `application/json`

## Security

### Basic

*Type* : basic

## Paths

**Get project details.**

```
GET /projects/{project_id}/
```

## Description

Get all project related information like Labels, SectionLabels (called Template).

## Parameters

Type	Name	Schema
Path	<b>project_id</b> <i>required</i>	string

## Responses

HTTP Code	Schema
200	<a href="#">Project</a>

## Tags

- projects

## Get a list of all documents in your project.

```
GET /projects/{project_id}/docs/
```

## Description

Use this endpoint to get meta information for all documents in your projects. You can use the `<i>created_at</i>` or `<i>id</i>` field to identify recently added documents. The default limit of returned documents for a request is 1000.

## Parameters

Type	Name	Description	Schema
Path	<b>project_id</b> <i>required</i>		string
Query	<b>limit</b> <i>optional</i>	Number of results to return per page.	integer

Type	Name	Description	Schema
Query	<b>offset</b> <i>optional</i>	The initial index from which to return the results.	integer

## Responses

HTTP Code	Schema
200	<a href="#">Response 200</a>

### Response 200

Name	Schema
<b>count</b> <i>required</i>	integer
<b>next</b> <i>optional</i>	string (uri)
<b>previous</b> <i>optional</i>	string (uri)
<b>results</b> <i>required</i>	< <a href="#">DocumentMeta</a> > array

## Tags

- projects

## Create a new annotation in a document.

```
POST /projects/{project_id}/docs/{doc_id}/annotations/
```

### Description

If you want to create annotations based on custom extraction results, set the following attributes in the post request: - label - start\_offset - end\_offset - accuracy - section\_label\_id

Use the project endpoint (/projects/{project\_id}/) to get the labels and section\_labels.

### Parameters

Type	Name	Schema
Path	<b>doc_id</b> <i>required</i>	string
Path	<b>project_id</b> <i>required</i>	string
Body	<b>data</b> <i>required</i>	<a href="#">SequenceAnnotation</a>

## Responses

HTTP Code	Schema
201	<a href="#">SequenceAnnotation</a>

## Tags

- projects

# Get all annotations for a document.

```
GET /projects/{project_id}/docs/{doc_id}/annotations/
```

## Description

Get all annotations for a document as a list.

## Parameters

Type	Name	Schema
Path	<b>doc_id</b> <i>required</i>	string
Path	<b>project_id</b> <i>required</i>	string

## Responses

HTTP Code	Schema
200	< <a href="#">SequenceAnnotation</a> > array

## Tags

- projects

## Get paragraphs for a document.

```
GET /projects/{project_id}/docs/{doc_id}/paragraph/
```

### Description

The return value represents the lines per paragraph per page. Therefore the return value is of type `List[List[Dict]]`.

### Parameters

Type	Name	Description	Schema
Path	<b>doc_id</b> <i>required</i>		string
Path	<b>project_id</b> <i>required</i>		string
Query	<b>height</b> <i>optional</i>	Force overwrite of dynamically calculated height used for paragraph splitting.	integer

### Responses

HTTP Code	Schema
200	No Content

## Tags

- projects

## Get segmentation for a document.

```
GET /projects/{project_id}/docs/{doc_id}/segmentation/
```

### Description

The return value represents the segmentation bounding boxes per page. Therefore the return value is of type `List[List[Dict]]`.

## Parameters

Type	Name	Schema
Path	<b>doc_id</b> <i>required</i>	string
Path	<b>project_id</b> <i>required</i>	string

## Responses

HTTP Code	Schema
200	No Content

## Tags

- projects

## Upload a new document.

POST /v2/docs/

## Description

By default the processing of a document is asynchronous. You can set a `callback_url` to get notified once the processing has finished or pull the results in a separate request using the `extraction_url` attribute.

Alternatively, you can make the processing synchronous by setting `sync=True`. Keep in mind that the processing can take several minutes for large documents.

## Parameters

Type	Name	Schema
FormData	<b>callback_url</b> <i>optional</i>	string (uri)
FormData	<b>data_file</b> <i>required</i>	file
FormData	<b>project</b> <i>required</i>	integer

Type	Name	Schema
FormData	sync <i>optional</i>	boolean

## Responses

HTTP Code	Schema
201	<a href="#">Response 201</a>

### Response 201

Name	Description	Schema
<b>callback_url</b> <i>optional</i>	<b>Maximal length : 200</b>	string (uri)
<b>data_file</b> <i>optional</i> <i>read-only</i>		string (uri)
<b>data_file_name</b> <i>optional</i> <i>read-only</i>	<b>Minimum length : 1</b>	string
<b>extraction_url</b> <i>optional</i> <i>read-only</i>		string
<b>id</b> <i>optional</i> <i>read-only</i>		integer
<b>project</b> <i>required</i>		integer
<b>sync</b> <i>optional</i>		boolean

## Consumes

- `multipart/form-data`
- `application/x-www-form-urlencoded`

## Tags

- docs

## Retrieve the extraction results for a document.

```
GET /v2/docs/{doc}/
```

### Description

Get all information for a document using its id. The extraction results are available once the processing has finished (status == [2, "Done"]).

### Parameters

Type	Name	Schema
Path	<b>doc</b> <i>required</i>	string

### Responses

HTTP Code	Schema
200	No Content

## Tags

- docs

## Give feedback to the extraction results of a document.

```
PATCH /v2/docs/{doc}/
```

### Description

You can give feedback by sending the adapted extraction result dictionary.

Example python code:



```

import requests
import json
from requests.auth import HTTPBasicAuth

url = f"https://app.konfuzio.com/api/v2/docs/{DOC_ID}/"
auth = HTTPBasicAuth(KONFUZIO_USER, KONFUZIO_PASSWORD)
data = requests.get(url=url, auth=auth).json()

# Mark Extraction as correct.
data['labels']['Bruttoszahweise']['extractions'][0]['correct'] = True

# Add new Extraction which has not been in result list.
data['labels']['Bruttoszahweise']['extractions'].append({'value': '123,45'})

r = requests.patch(url = url, data=json.dumps(data), auth=auth, headers={'Content-
Type': 'application/json'},)

```

## Parameters

Type	Name	Schema
Path	<b>doc</b> <i>required</i>	string

## Responses

HTTP Code	Schema
200	No Content

## Tags

- docs

# Definitions

## DocumentMeta

Name	Description	Schema
<b>callback_status_code</b> <i>optional</i>	<b>Minimum value</b> : -2147483648 <b>Maximum value</b> : 2147483647	integer

Name	Description	Schema
<b>callback_url</b> <i>optional</i>	<b>Maximal length : 200</b>	string (uri)
<b>data_file</b> <i>optional</i> <i>read-only</i>		string (uri)
<b>data_file_name</b> <i>required</i>	<b>Length : 1 - 1024</b>	string
<b>data_file_producer</b> <i>optional</i>	<b>Maximal length : 1024</b>	string
<b>dataset_status</b> <i>optional</i>		integer
<b>extraction_time</b> <i>optional</i>		number
<b>file_url</b> <i>optional</i> <i>read-only</i>		string
<b>id</b> <i>optional</i> <i>read-only</i>		integer
<b>number_of_pages</b> <i>optional</i>	<b>Minimum value : -2147483648</b> <b>Maximum value : 2147483647</b>	integer
<b>ocr_time</b> <i>optional</i>		number
<b>status</b> <i>optional</i> <i>read-only</i>		string
<b>updated_at</b> <i>optional</i> <i>read-only</i>		string (date-time)

Name	Description	Schema
<b>use_fallback_ocr</b> <i>optional</i>		boolean
<b>workflow_end_time</b> <i>optional</i>		number
<b>workflow_start_time</b> <i>optional</i>		number

## Label

Name	Description	Schema
<b>description</b> <i>optional</i>		string
<b>get_data_type_display</b> <i>optional</i> <i>read-only</i>		string
<b>has_multiple_top_candidates</b> <i>optional</i>	If this is set, (1) annotations with this label are combined into a single cell in the csv export, (2) multiple extractions with this label are shown on the dashboard (instead of the most likely one), (3) in the evaluation every annotation of the ground truth needs to be matched (instead of only one).	boolean
<b>id</b> <i>optional</i> <i>read-only</i>		integer
<b>shortcut</b> <i>optional</i> <i>read-only</i>		string
<b>text</b> <i>required</i>	<b>Length : 1 - 100</b>	string

Name	Description	Schema
<b>text_clean</b> <i>optional</i> <i>read-only</i>	<b>Minimum length : 1</b>	string
<b>threshold</b> <i>optional</i>		number
<b>token_full_replacement</b> <i>optional</i>		boolean
<b>token_number_replacement</b> <i>optional</i>		boolean
<b>token_whitespace_replacement</b> <i>optional</i>		boolean

## Project

Name	Description	Schema
<b>enable_api</b> <i>optional</i>	Decide if this project can be accessed via API. API access is necessary for uploading new documents.	boolean
<b>id</b> <i>optional</i> <i>read-only</i>		integer
<b>labels</b> <i>optional</i> <i>read-only</i>		< Label > array
<b>name</b> <i>required</i>	<b>Length : 1 - 100</b>	string
<b>priority_ocr</b> <i>optional</i>		boolean

Name	Description	Schema
<b>section_labels</b> <i>optional</i> <i>read-only</i>		< <a href="#">SectionLabel</a> > array
<b>storage_name</b> <i>optional</i>	<b>Maximal length : 100</b>	string

## SectionLabel

Name	Description	Schema
<b>has_multiple_sections</b> <i>optional</i>	Decide if a section with this template can occur multiple times in a document. For example, an invoice can have multiple positions. A default template cannot have multiple sections.	boolean
<b>id</b> <i>optional</i> <i>read-only</i>		integer
<b>is_default</b> <i>optional</i>	Default templates refer to the type of a document, e.g. invoice, contract, payslip, account statement.	boolean
<b>labels</b> <i>optional</i> <i>read-only</i>		< integer > array
<b>name</b> <i>required</i>	<b>Length : 1 - 25</b>	string
<b>name_clean</b> <i>optional</i> <i>read-only</i>	<b>Minimum length : 1</b>	string

## SequenceAnnotation

Name	Description	Schema
<b>accuracy</b> <i>optional</i>		number
<b>bbox</b> <i>optional</i>		string

Name	Description	Schema
<b>bboxes</b> <i>optional</i> <i>read-only</i>		string
<b>created_by</b> <i>optional</i>		integer
<b>custom_bboxes</b> <i>optional</i>		object
<b>custom_offset_string</b> <i>optional</i> <i>read-only</i>		string
<b>end_offset</b> <i>optional</i>	<b>Minimum value :</b> -2147483648 <b>Maximum value :</b> 2147483647	integer
<b>get_created_by</b> <i>optional</i> <i>read-only</i>		string
<b>get_revised_by</b> <i>optional</i> <i>read-only</i>		string
<b>id</b> <i>optional</i> <i>read-only</i>		integer
<b>is_correct</b> <i>optional</i>		boolean
<b>label</b> <i>required</i>		integer
<b>label_data_type</b> <i>optional</i> <i>read-only</i>		string

Name	Description	Schema
<b>label_text</b> <i>optional</i> <i>read-only</i>		string
<b>label_threshold</b> <i>optional</i> <i>read-only</i>		string
<b>normalized</b> <i>optional</i> <i>read-only</i>		string
<b>offset_string</b> <i>optional</i>		string
<b>offset_string_original</b> <i>optional</i>		string
<b>page_number</b> <i>optional</i>		integer
<b>revised</b> <i>optional</i>		boolean
<b>revised_by</b> <i>optional</i>		integer
<b>section</b> <i>optional</i>		integer
<b>section_label_id</b> <i>optional</i>		integer
<b>section_label_text</b> <i>optional</i> <i>read-only</i>		string
<b>selection_box</b> <i>optional</i>		object

Name	Description	Schema
<b>start_offset</b> <i>optional</i>	<b>Minimum value :</b> -2147483648 <b>Maximum value :</b> 2147483647	integer
<b>translated_string</b> <i>optional</i>		string