

- **401 Unauthorized (Authorization Error)**

```
JSON
{
  "error": "access_denied",
  "error_description": "Unauthorized"
}
```

1.1. Document Extraction

POST /extract/

- **Description:** Processes a PDF file and extracts structured data based on the specified document type.
- **Authentication:** * **Header:** `Authorization: Bearer <access_token>`
 - **Required:** Yes. Use the token obtained from the /oauth/token endpoint.
- **Request Body (Form Data):**
 - **document_type:** (String, **Required**) - The type of document to be processed. Must be one of the `DocumentType` enum values.
 - Example: `export_declaration`, `hbl`, `commercial_invoice`
 - **file:** (File, **Required**) - The PDF document file to be processed.
- **Responses:**
 - **200 OK (Success)**

```
JSON

{
  "time": 11.734270811080933,
  "message": "The document processed successfully",
  "identifier_page": 1,
  "extractor_page": 1,
  "process_id": "3a0a44f5-acef-49f2-9421-ebfba387169b",
  "response": [
    {
      "consigneeAddress": {
```

```

        "bbox": [0.102, 0.05, 0.450, 0.120],
        "page": 1,
        "value": "My Street and number"
    },
    "invoiceAmount": {
        "bbox": [0.7055, 0.5708, 0.8756, 0.5842],
        "page": 1,
        "value": "830.0"
    },
    "currency": {
        "bbox": [0.880, 0.570, 0.920, 0.584],
        "page": 1,
        "value": "GBP"
    },
    "skuData": [
        {
            "itemNo": {
                "bbox": [0.5, 0.1, 0.52, 0.12],
                "page": 1,
                "value": 1
            },
            "descriptionOfGoods": {
                "bbox": [0.5, 0.2, 0.55, 0.4],
                "page": 1,
                "value": "NEW MERCEDES BENZ VAN"
            }
        }
    ]
}

```

- **401 Unauthorized (Authorization Error)**

```

JSON
{
  "msg": "Access Denied",
  "remark": "A JWT is required as a Bearer token.",
  "decision_id": "auth-service-internal"
}

```

- **403 Forbidden**

JSON

```
{
  "detail": "You are not authorized to process document type: <document_type>"
}
```

1.2. Combined Classification and Extraction

POST `/classify_extract/`

- **Description:** First classifies the PDF document, then splits the document based on the identified document types and performs extraction on each classified section.
- **Authentication:** * **Header:** `Authorization: Bearer <access_token>`
 - **Required:** Yes. Use the token obtained from the `/oauth/token` endpoint.
- **Request Body (Form Data):**
 - **file:** (File, **Required**) - The PDF document file to be processed.
- **Responses:**
 - **200 OK (Success)**

JSON

```
{
  "time": 11.734270811080933,
  "message": "The document processed successfully",
  "identifier_page": 1,
  "extractor_page": 1,
  "process_id": "2a0a71f5-acef-49f2-9421-ebfba387169b",
  "response": [
    {
      "consigneeAddress": {
        "bbox": [0.102, 0.05, 0.450, 0.120],
        "page": 1,
        "value": "My Street and number"
      },
      "invoiceAmount": {
        "bbox": [0.7055, 0.5708, 0.8756, 0.5842],
        "page": 1,
      }
    }
  ]
}
```

```
    "value": "830.0"
  },
  "currency": {
    "bbox": [0.880, 0.570, 0.920, 0.584],
    "page": 1,
    "value": "GBP"
  },
  "skuData": [
    {
      "itemNo": {
        "bbox": [0.5, 0.1, 0.52, 0.12],
        "page": 1,
        "value": 1
      },
      "descriptionOfGoods": {
        "bbox": [0.5, 0.2, 0.55, 0.4],
        "page": 1,
        "value": "NEW MERCEDES BENZ VAN"
      }
    }
  ]
}
```

- **401 Unauthorized (Authorization Error)**

JSON

```
{
  "msg": "Access Denied",
  "remark": "A JWT is required as a Bearer token.",
  "decision_id": "auth-service-internal"
}
```

1.3. Async Classification and Extraction

POST `/classify_extract_async/`

- **Description:** Submits a PDF file for asynchronous classification and data extraction. The API returns a `process_id` that you can use to check the status and retrieve the results later.
- **Authentication:** * **Header:** `Authorization: Bearer <access_token>`
 - **Required:** Yes. Use the token obtained from the `/oauth/token` endpoint.
- **Request Body (Form Data):**
 - **file:** (File, Required) - The PDF document file to be processed.
- **Responses:**
 - **200 OK (Success)**

```
JSON
{
  "time": "2.4480321",
  "message": "The document queued successfully.",
  "process_id": "3185f3b6-9673-87fa-38c8a94321d"
}
```

- **401 Unauthorized (Authorization Error)**

```
JSON
{
  "msg": "Access Denied",
  "remark": "A JWT is required as a Bearer token.",
  "decision_id": "auth-service-internal"
}
```

POST `/extract_async/`

- **Description:** Submits a PDF or Image file for extraction using a specific document type. The API returns a `process_id` immediately, and the processing happens in the background. Recommended for high-volume or multi-page documents.
- **Authentication:** * **Header:** `Authorization: Bearer <access_token>`

- **Request Body (Form Data):**
 - **document_type:** (String, **Required**) - One of the supported **DocumentType** values.
 - **file:** (File, **Required**) - The document file to be processed.
- **Responses:**
 - **200 OK (Success)**

JSON

```
{
  "time": "2.4480321",
  "message": "The document queued successfully.",
  "process_id": "3185f3b6-9673-87fa-38c8a94321d"
}
```

- **401 Unauthorized (Authorization Error)**

JSON

```
{
  "msg": "Access Denied",
  "remark": "A JWT is required as a Bearer token.",
  "decision_id": "auth-service-internal"
}
```

1.4. Status and Results

GET `/get_status/`

- **Description:** Checks the current status of a document processing job using the provided `process_id`.
- **Authentication:** * **Header:** `Authorization: Bearer <access_token>`
 - **Required:** Yes. Use the token obtained from the `/oauth/token` endpoint.
- **Request Params:**
 - `process_id`: (String, **Required**). The unique identifier for the processing job.
- **Responses:**
 - **200 OK (Success)**

JSON

```
{
  "time": "2.4480321",
  "message": "The document processed sucessfully.",
  "process_id": "3185f3b6-9673-87fa-38c8a94321d",
  "response": "Success"
}
```

- **400 Bad Request**

JSON

```
{
  "time": "2.4480321",
  "message": "The process id not found!",
  "process_id": "3185f3b6-9673-87fa-38c8a94321d",
  "response": null
}
```

- **401 Unauthorized (Authorization Error)**

JSON

```
{
  "msg": "Access Denied",
  "remark": "A JWT is required as a Bearer token.",
  "decision_id": "auth-service-internal"
}
```

GET /get_result/

- **Description:** Retrieves the final extraction results for a completed document processing job.
- **Authentication:** *Header: Authorization: Bearer <access_token>
 - **Required:** Yes. Use the token obtained from the /oauth/token endpoint.
- **Request Params:**
 - **process_id:**(String, **Required**). The unique identifier for the processing job.
 - **bbox:** (Boolean, **Optional**): Include bounding box coordinates (*The extraction and exact location might not match, Extraction is more reliable in this case!*) (Default: **true**, for faster response set it to **false**)
- **Responses:**
 - **200 OK (Success)**

JSON

```
{
  "time": "2.4480321",
  "message": "The document processed successfully.",
  "extractor_page": 1,
  "process_id": "3185f3b6-9673-87fa-38c8a94321d",
  "response": [
    {
      "consigneeAddress": {
        "bbox": [0.102, 0.05, 0.450, 0.120],
        "page": 1,
        "value": "My Street and number"
      },
      "invoiceAmount": {
        "bbox": [0.7055, 0.5708, 0.8756, 0.5842],
        "page": 1,
      }
    }
  ]
}
```

```

        "value": "830.0"
    },
    "currency": {
        "bbox": [0.880, 0.570, 0.920, 0.584],
        "page": 1,
        "value": "GBP"
    },
    "skuData": [
        {
            "itemNo": {
                "bbox": [0.5, 0.1, 0.52, 0.12],
                "page": 1,
                "value": 1
            },
            "descriptionOfGoods": {
                "bbox": [0.5, 0.2, 0.55, 0.4],
                "page": 1,
                "value": "NEW MERCEDES BENZ VAN"
            }
        }
    ]
}

```

- **400 Bad Request**

JSON

```

{
    "time": "2.4480321",
    "message": "The process id not found!",
    "extractor_page": null,
    "process_id": "3185f3b6-9673-87fa-38c8a94321d",
    "response": null
}

```

1.5. Utility Endpoints

GET /get_document_types/

- **Description:** Returns a list of all valid strings supported for the `document_type` parameter.

GET /get_file/

- **Description:** Downloads the binary file (original or processed/rotated) associated with a process.
- **Params:**
 - `process_id`: (String, **Required**).
 - `extension`: (String, **Required**) - e.g., pdf, png, jpg.