SKOPOS' SCANNING SERVICE DOCUMENTATION

[View the Github Repository here!](#)

Overview

Skopos' scanning service provides endpoints for users to perform PDF scans with the hopes of receiving a text output extracted from the pdf. The service provides a secure API using OAuth2 authentication with client credential grant type.

Endpoints

1. Register Client

`/register-client`

2. Get Access Token

`/oauth/token`

3. Upgrade to Exclusive Tier

`/upgrade-to-exclusive`

4. Extract Text from PDF (Freemium)

`/extract-text/`

5. Extract Text from PDF with OCR (Exclusive)

`/extract-text-ocr/`

6. Search for Keywords

`/search-text/`

Authentication Flow

The service provided utilizes Oauth2 authentication with client-credentials grant type, here are the following steps to use our API:

1. Register the client and obtain the client_id and client_secret.
2. Using the received credentials, request for a token to gain a temporary bearer token access.
3. Include the token access into the header of each endpoint request.

DETAILED API REFERENCE

1. Register Client

- **Endpoint:** `/register-client`
- **HTTP Method:** `POST`
- **Description:** Creates a new client and returns API credentials. The `client_secret` is only provided once in this response.
- **Request Header:** None needed.
- **Request Body:** (`application/json`)

  ```
  {
    "client_name": "string",
    "email": "string",
    "redirect_uri": "string"
  }
  ```

- **Response Body:** (`application/json`)

```
{
  "client_id": "string",
  "client_secret": "string",
  "client_name": "string",
  "email": "string",
  "redirect_uri": "string",
  "tier": "freemium"
}
```

2. Get Access Token

- **Endpoint:** `/oauth/token`
- **HTTP Method:** `POST`
- **Description:** Exchanges client credentials for a temporary bearer token.
- **Request Header:** None needed.
- **Request Body:** (`application/x-www-form-urlencoded`)
  - `grant_type` : "client_credentials"
  - `client_id` : "your_client_id"
  - `client_secret` : "your_client_secret"
- **Response Body:** (`application/json`)

```
{
  "access_token": "string",
  "token_type": "bearer"
}
```

3. Upgrade to Exclusive Tier

- **Endpoint:** `/upgrade-to-exclusive`
- **HTTP Method:** `POST`
- **Description:** Upgrades the authenticated client's account tier to `exclusive` .
- **Request Header:**
  - `Authorization` : "Bearer YOUR_ACCESS_TOKEN"
- **Request Body:** None.
- **Response Body:** (`application/json`)

```
{
  "message": "Upgrade successful!",
  "client": {
    "client_name": "string",
    "email": "string",
    "redirect_uri": "string",
    "client_id": "string",
    "tier": "exclusive"
  }
}
```

4. Extract Text from PDF (Freemium)

- **Endpoint:** `/extract-text/`
- **HTTP Method:** `POST`
- **Description:** Extracts text directly from a PDF file. Available to all authenticated users with a 5MB file limit.
- **Request Header:**
  - `Authorization` : "Bearer YOUR_ACCESS_TOKEN"
- **Request Body:** (`multipart/form-data`)
  - `file` : The PDF file to be processed.
- **Response Body:** (`application/json`)

```
{
  "filename": "string",
  "text": "string",
  "authorized_client": "string",
  "tier_access": "freemium"
}
```

5. Extract Text from PDF with OCR (Exclusive)

- **Endpoint:** `/extract-text-ocr/`
- **HTTP Method:** `POST`
- **Description:** Extracts text from a PDF using OCR, ideal for scanned documents. Requires `exclusive` tier and has a 50MB file limit.
- **Request Header:**
  - `Authorization` : "Bearer YOUR_ACCESS_TOKEN"
- **Request Body:** (`multipart/form-data`)
  - `file` : The PDF file to be processed.
- **Response Body:** (`application/json`)

```
{
  "filename": "string",
  "ocr_text": "string",
  "authorized_client": "string",
  "tier_access": "exclusive"
}
```

---

6. Search for Keywords

- **Endpoint:** `/search-text/`
- **HTTP Method:** `POST`
- **Description:** Searches for multiple keywords in an uploaded text file. Available to all authenticated users.
- **Request Header:**
  - `Authorization` : "Bearer YOUR_ACCESS_TOKEN"
- **Request Body:** (`multipart/form-data`)
  - `file` : The text file to search within.
  - `keywords` : A keyword to search for (this field can be repeated for multiple keywords).
- **Response Body:** (`application/json`)

```
{
  "filename": "string",
  "found_keywords": [
    {
      "keyword": "string",
      "start_index": 0,
      "end_index": 0
    }
  ],
  "authorized_client": "string",
  "tier_access": "freemium"
}
```