

# Web Scraping Task: Israeli Supreme Court Decisions - Document Download

## Objective

Create a Python script to scrape and **download court decision documents** (DOCX/PDF files) from the Israeli Supreme Court website for a single day using HTTP requests.

## Target URL

<https://supreme.court.gov.il/>

## Requirements

### 1. Core Functionality

- Search for court decisions for a **single day** (date must be configurable)
- **Download the actual decision documents** (DOCX or PDF files)
- Use **HTTP requests only** (no browser automation)
- Save documents with meaningful filenames
- Create a metadata JSON file listing all downloads

### 2. Search Parameters to Support

Based on the site's filtering capabilities, implement support for:

- **Date range** (required) - start date and end date
- **Decision type** (סוג החלטה)
- **Case type** (סוג תיק)
- **Keywords** (מילים Chi'afot)
- Any other filters available on the search form

### 3. Data to Extract and Download

For each decision:

- **Download the document file** (DOCX or PDF)
- Extract metadata:
  - Case number (מספר תיק)

- תאריך החלטה (Decision date)
- בעלי דין (Parties involved)
- Type of document (PDF/DOCX)
- File size
- Download URL

## 4. File Naming Convention

Use a clear naming pattern:

{case\_number}\_{date}\_{sequential\_number}.{extension}

Example: 001\_2025-11-23\_25\_1234\_עבג.pdf

Since Hebrew characters might cause issues, consider sanitizing:

evag\_1234\_25\_2025-11-23\_001.pdf

## 5. Output Structure

```
output/
  └── documents/
      ├── case_001_2025-11-23.pdf
      ├── case_002_2025-11-23.docx
      └── case_003_2025-11-23.pdf
  └── metadata.json
  └── download_log.txt
```

## 6. Metadata JSON Format

```
{
  "search_date": "2025-11-23",
  "download_timestamp": "2025-11-23T10:30:00",
  "total_documents": 15,
  "successful_downloads": 14,
  "failed_downloads": 1,
  "documents": [
    {
      "case_number": "1234/25",
      "decision_date": "2025-11-23",
      "parties": "ישראל ישראלי נ' מדינת ישראל",
      "document_type": "pdf",
      "file_size_bytes": 245760,
```

```
        "filename": "case_001_2025-11-23.pdf",
        "download_url": "https://supreme.court.gov.il/...",
        "download_status": "success"
    }
]
}
```

## Deliverables

1. `scraper.py` - Main Python script
2. `requirements.txt` - Python dependencies
3. `README.md` - Detailed usage instructions
4. `config.example.json` - Example configuration file (optional)

## Example Usage

```
# Download all decisions from a specific date
python scraper.py --date 2025-11-23 --output-dir ./output
```

```
# With additional filters
python scraper.py --date 2025-11-23 --decision-type "ל'ת_709" --output-dir ./output
```