



PDS Data Optimization User Manual

Victor Guara
Brian Kwon
Kevin Ng
Mangat Toor
Lisa Zhu

May. 30th, 2024

Table of Contents

Introduction.....	1
System Requirements.....	1
Installation.....	1
Download.....	1
Setup.....	2
Start Application.....	2
Usage.....	4
Consequential Amendments.....	4
Generate Consequential Amendments.....	6
Generate Sections.....	9
Digitization of District Schedule.....	10
Updating Documents.....	10
Exit Application.....	11
Restart application.....	12
Exclude Downloaded PDF Files.....	13

Introduction

This application efficiently identifies and catalogs terms in PDF documents, scanning around 500 publicly available PDF files from the City of Vancouver website for user-defined terms and related terms. It then outputs an Excel file containing detailed information about each occurrence, including document titles, section numbers, and reference paragraphs. Moreover, it proposes updates to documents based on user-defined prompts, processing amendments with suggested updates. Additionally, the application extracts and organizes data from PDF documents for database integration, fetching selected files, and allowing users to define data fields.

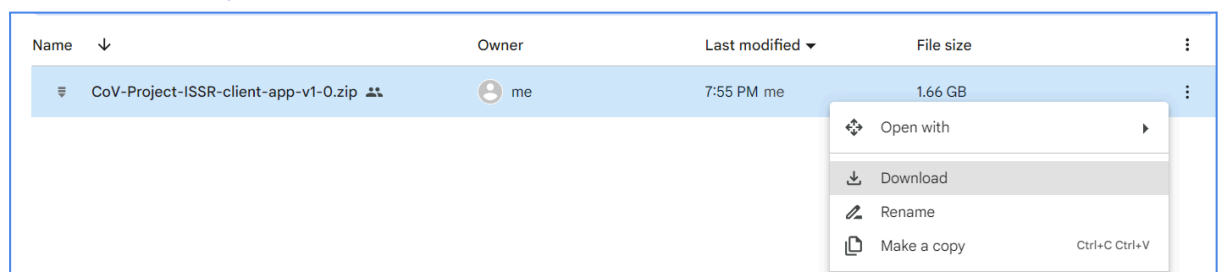
System Requirements

Operating System	Windows
Web Browser	Google Chrome, Mozilla Firefox, Microsoft Edge

Installation

Download

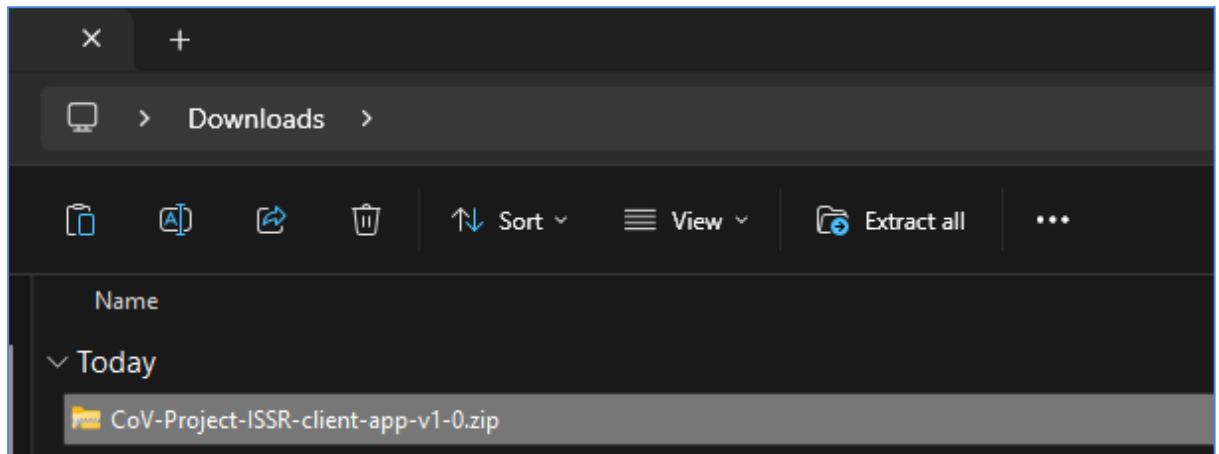
1. User can download the application zip file from our Google Drive with following link (may change after 31 May 2024):
https://drive.google.com/file/d/1Jg_SKR_NK9GSSHa2GEgOKNTvWXn4QRkr/view?usp=drive_link



Note: The file is approximately 1.7 GB and may take up to 5-10 minutes to download.

Setup

2. Unzip the CoV-Project-ISSR-client-app-v(x).zip file in any non-restricted folder of the drive C.



Note: Do not choose drive F or H

3. The following list of files should appear after unzipping.

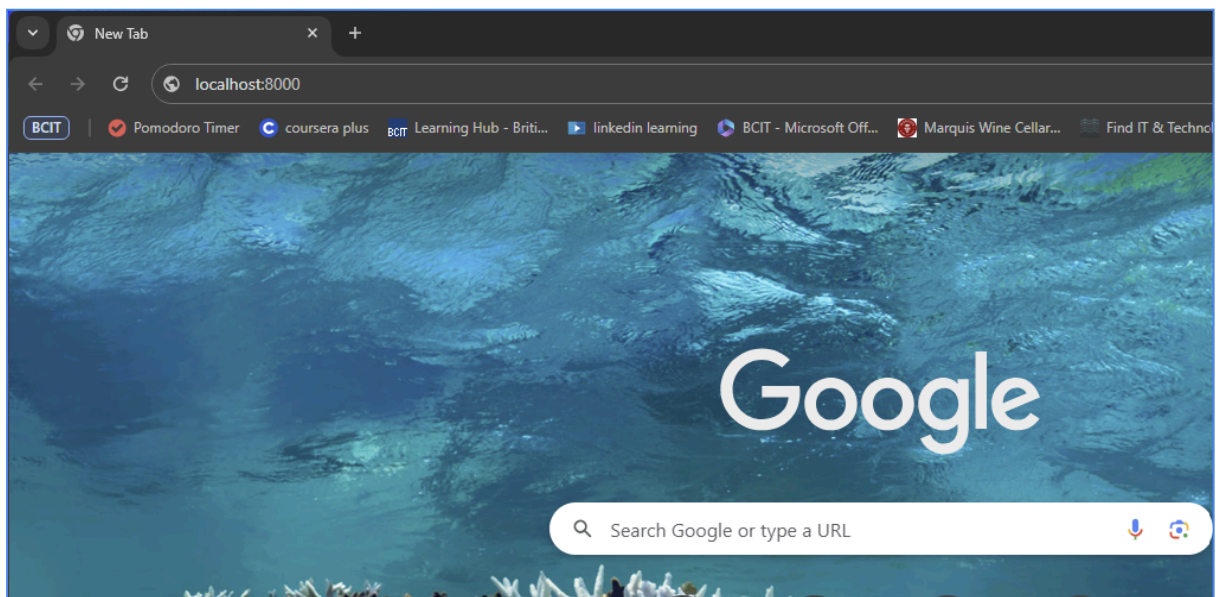
Name	Date modified	Type	Size
downloaded_pdfs	28/5/2024 3:10 pm	File folder	
credential.ini	28/5/2024 1:22 pm	Configuration sett...	1 KB
development.ini	27/5/2024 6:16 pm	Configuration sett...	1 KB
doc_type.json	25/5/2024 12:58 am	JSON File	253 KB
processed_final.json	22/5/2024 10:51 pm	JSON File	25,689 KB
server.exe	29/5/2024 12:16 am	Application	439,489 KB

Start Application

4. Double-click the server.exe file to start the application. A window will open.
5. After 3-5 minutes, the following screen will show. Indicating the application is ready to access.

```
C:\Users\kevin\Downloads\te: X + v
* Serving Flask app 'server'
* Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
* Running on http://127.0.0.1:8000
Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 180-185-503
```

6. From any browser, type localhost:8000 or 127.0.0.1:8000 and press Enter to access the application.



- When you see the screen below, the application has successfully started.

The screenshot shows the City of Vancouver Land Use Document Search Tool interface. The header includes the City of Vancouver logo, the title "PDS Data Optimization Proof of Concept Land Use Document Search Tool for Consequential Amendments and Digitizing District Schedule", and two buttons: "Manual" and "Update Files". Below the header, there is a left-hand sidebar with two options: "Consequential Amendments" (highlighted with a blue background) and "Digitization of District Schedule". The main content area contains an information box with a blue header and a message: "Please update the files by clicking the 'Update Files' button on the top right and clicking 'update'." Below this, there is a search bar with the placeholder text "Search keyword(s)..." and a button "Search keyword(s)...". Under the search bar, there is a section titled "Select document(s) and/or document type(s) from the list:" with a dropdown menu showing "All X" and "Filter tags". At the bottom of the main content area, there are two checkboxes: "Generate Consequential Amendments" and "Generate Sections", both of which are unchecked. A "Search" button is located at the bottom right of the main content area.

Note: Do not close the black screen application Windows until you decide to exit the application.

Usage

Consequential Amendments

- Click on “Consequential Amendments” on the left hand navbar.

This screenshot is identical to the one above, showing the City of Vancouver Land Use Document Search Tool interface. The only difference is that the "Consequential Amendments" option in the left-hand sidebar is now highlighted with a red border, indicating it is the selected option.

2. Enter keyword(s) into the search box by pressing “Enter”, or separated by a comma.

Type in a keyword(s) below, press 'enter' or ',' to separate keywords:

parking X Search keyword(s)...

X

Select document(s) and/or document type(s) from the list:

All X Filter tags

X

Generate Consequential Amendments ☐

Generate Sections ☐

Search

3. Select filters as needed. The application will search all documents by default.

Type in a keyword(s) below, press 'enter' or ',' to separate keywords:

parking X Search keyword(s)...

X

Select document(s) and/or document type(s) from the list:

All X Filter tags

X

Categories

Zoning and Development By-law
Official development plan by-laws
Policies
Guidelines
City-owned or leased facility guidelines

4. Click on “Search”.

Type in a keyword(s) below, press 'enter' or ',' to separate keywords:

parking X Search keyword(s)...

X

Select document(s) and/or document type(s) from the list:

All X Filter tags

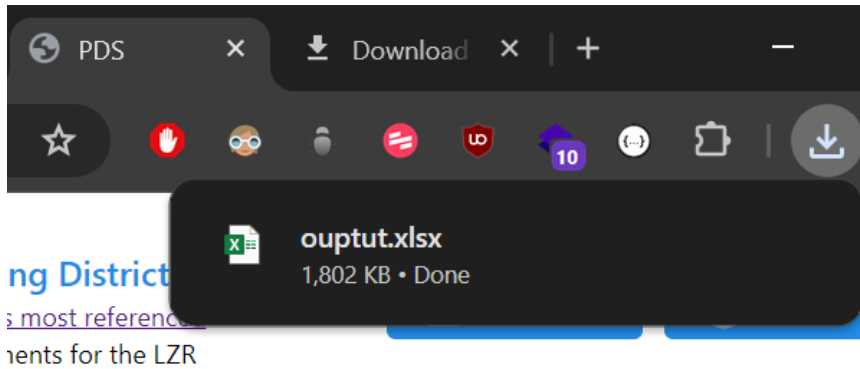
X

Generate Consequential Amendments ☐

Generate Sections ☐

Search

- After a moment, an Excel file will be generated.



- Below is the Excel output for the previous configurations.

Docume	File name	Title	AI Title	Section #	ection Title	Search Term	Link	Page Numbr	Reference	Used amen	Rationale
1	Guideline: B001.pdf	900 Burrar	No AI Title	Unknown	Unknown	parking	https://gu1	50	City of No	Amenc	No Rationale
2	Guideline: B001.pdf	900 Burrar	No AI Title	Unknown	Unknown	parking	https://gu6		City of Var	No	Amenc No Rationale
3	Guideline: B001.pdf	900 Burrar	No AI Title	Unknown	Unknown	parking	https://gu7		City of Var	No	Amenc No Rationale
4	Guideline: B002.pdf	1601 Wes!	No AI Title	Unknown	Unknown	parking	https://gu1		Page 1 A	No	Amenc No Rationale
5	Guideline: B002.pdf	1601 Wes!	No AI Title	Unknown	Unknown	parking	https://gu10		3.10 Parki	No	Amenc No Rationale
6	Guideline: B002.pdf	1601 Wes!	No AI Title	Unknown	Unknown	parking	https://gu10		Commerci	No	Amenc No Rationale
7	Guideline: B002.pdf	1601 Wes!	No AI Title	Unknown	Unknown	parking	https://gu18		City of Va	No	Amenc No Rationale
8	Guideline: B002.pdf	1601 Wes!	No AI Title	Unknown	Unknown	parking	https://gu2		Page 1 A	No	Amenc No Rationale
9	Guideline: B004.pdf	201 Burrar	No AI Title	Unknown	Unknown	parking	https://gu10		The lower	No	Amenc No Rationale
10	Guideline: B004.pdf	201 Burrar	No AI Title	Unknown	Unknown	parking	https://gu10		3.8 Parkin	No	Amenc No Rationale
11	Guideline: B004.pdf	201 Burrar	No AI Title	Unknown	Unknown	parking	https://gu10		Garbage s	No	Amenc No Rationale
12	Guideline: B004.pdf	201 Burrar	No AI Title	Unknown	Unknown	parking	https://gu13		City of Va	No	Amenc No Rationale
13	Guideline: B004.pdf	201 Burrar	No AI Title	Unknown	Unknown	parking	https://gu17		City of Va	No	Amenc No Rationale
14	Guideline: B004.pdf	201 Burrar	No AI Title	Unknown	Unknown	parking	https://gu22		(a) Portior	No	Amenc No Rationale
15	Guideline: B004.pdf	201 Burrar	No AI Title	Unknown	Unknown	parking	https://gu28		2A (Burrar	No	Amenc No Rationale
16	Bulletins bulletin-ar	Animal Gr	No AI Title	Unknown	Unknown	parking	https://by1		Backgrou	No	Amenc No Rationale
17	Bulletins bulletin-br	Building G	No AI Title	Unknown	Unknown	parking	https://by2		Requirem	No	Amenc No Rationale
18	Bulletins bulletin-cc	Communi	No AI Title	Unknown	Unknown	parking	https://by1		Should the	No	Amenc No Rationale
19	Bulletins bulletin-cc	Communi	No AI Title	Unknown	Unknown	parking	https://by2		When mal	No	Amenc No Rationale
20	Bulletins bulletin-di	Developm	No AI Title	Unknown	Unknown	parking	https://by5		DCIs appl	No	Amenc No Rationale
21	Bulletins bulletin-di	Developm	No AI Title	Unknown	Unknown	parking	https://by7		There are	No	Amenc No Rationale
22	Bulletins bulletin-er	Emergenc	No AI Title	Unknown	Unknown	parking	https://by1		(a) the cor	No	Amenc No Rationale
23	Bulletins bulletin-fl	Calculatio	No AI Title	Unknown	Unknown	parking	https://by3		Areas use	No	Amenc No Rationale
24	Bulletins										

Note: For section number/title, “Unknown” indicates an error occurred, while “Not found” indicates the AI responded, but the section does not exist.

Generate Consequential Amendments

- To generate amendments, click on the “Generate Consequential Amendments” checkbox.

Type in a keyword(s) below, press 'enter' or ',' to separate keywords:

 parking X Search keyword(s)... X

Select document(s) and/or document type(s) from the list:


 All X Filter tags X

Generate Consequential Amendments ☐

Generate Sections ☐

Search

8. Enter the AI prompt for changes needed. Then, click on “Search”.

Generate Consequential Amendments 

Generate Sections ☐

Please type your AI prompt to identify regulations in need of change and to propose a consequential amendment.

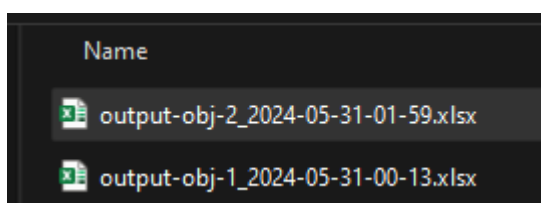
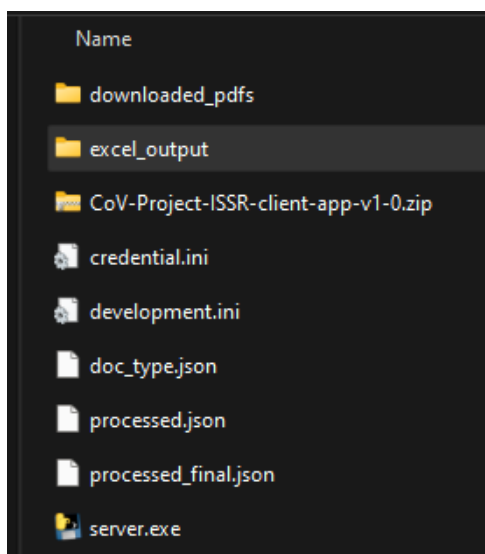
Example: If there is any mention of parking space size or the number of parking spots, replace it with 'See parking by-law'. Do not replace mentions of parking access or location. Do not replace if the parking by-law is already mentioned.

Search

9. After the process is finished. You can find the file inside the “excel_output” folder from the application directory

Search

File is created in folder excel_output with filename output-obj-2_2024-05-31-01-59.xlsx



Note: The output excel files for **Generate Consequential Amendments**, **Generate Sections** and **Digitization of District Schedule** are all stored in excel_output folder

Generate Sections

10. Enable the “Generate Sections” checkbox to generate section numbers/titles.

Type in a keyword(s) below, press 'enter' or ',' to separate keywords:

Select document(s) and/or document type(s) from the list:

Generate Consequential Amendments ☐

Generate Sections ☐

Search

11. Click on “Search”
12. File will be output to the excel_output folder

Note: For section number/title, there will be a 60 second timeout every 5 search results. For amendments, there will be a 60 second timeout every 20 search results. This is due to the free AI model limitations.

Note: Enabling this checkbox will increase the overall process time.

Digitization of District Schedule

1. Click on “Digitization of District Schedule” on the left hand navbar.

CITY OF VANCOUVER

PDS Data Optimization Proof of Concept

Land Use Document Search Tool for Consequential Amendments and Digitizing District Schedule

This application searches for keywords in documents at [Zoning and land use](#), and [Vancouver's most referenced bylaws](#). It can also create amendments based on keyword searches and digitize land use documents for the LZR database. Please refer to the [manual](#) for more information.

[Manual](#) [Update Files](#)

Consequential Amendments

Digitization of District Schedule

Please update the files by clicking the 'Update Files' button on the top right and clicking 'update'.

Select document(s) and/or document type(s) from the list:

Filter tags

Search

2. Apply filters as necessary, then click on “Search”. The Excel file will be output to the excel_output folder when the process finishes.

Select document(s) and/or document type(s) from the list:

Filter tags

Categories

- All
- Zoning and Development By-law
- Official development plan by-laws
- Policies
- Guidelines

Updating Documents

1. To update documents, click on the "Update Files" button in the top right corner.

CITY OF VANCOUVER

PDS Data Optimization Proof of Concept

Land Use Document Search Tool for Consequential Amendments and Digitizing District Schedule

This application searches for keywords in documents at [Zoning and land use](#), and [Vancouver's most referenced bylaws](#). It can also create amendments based on keyword searches and digitize land use documents for the LZR database. Please refer to the [manual](#) for more information.

[Manual](#) [Update Files](#)

Consequential Amendments

Digitization of District Schedule

Please update the files by clicking the 'Update Files' button on the top right and clicking 'update'.

Type in a keyword(s) below, press 'enter' or ';' to separate keywords:

Search keyword(s)...

Select document(s) and/or document type(s) from the list:

☐ All ☐ Filter tags

Generate Consequential Amendments ☐

Generate Sections ☐

Search

2. A popup window will appear. Click on "Update".

Status: Idle

Total Files: 590

2024-05-22 13:07:35

Update

3. The same popup window will show the update progress.

Status: Updating

Total Files: 590

2024-05-22 13:07:35



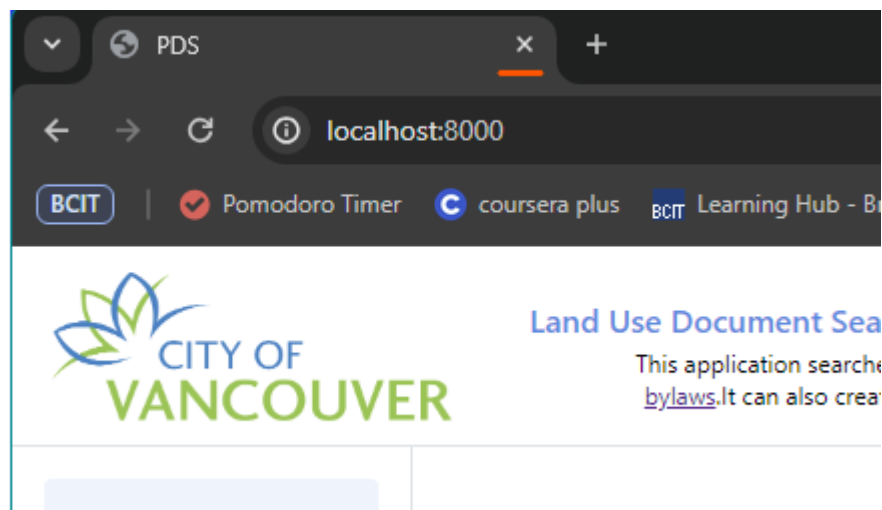
Note: Updating files may take 5-10 minutes to complete.

Exit Application

1. Click the close button (X) on the top right-handed corner to terminate the program. The application is now completely stopped.

```
C:\Users\kevin\Downloads\te: X + v
* Running on http://127.0.0.1:8000
Press CTRL+C to quit
* Restarting with stat
* Debugger is active!
* Debugger PIN: 180-185-503
[2024-05-29 00:38:06,263] INFO in server: A request is received for loading home page
127.0.0.1 - - [29/May/2024 00:38:06] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:06] "GET /static/css/c196650da8e3275d.css HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:06] "GET /static/css/4c8d1ee479f99e2d.css HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:06] "GET /static/city_vancouver_logo.png?w=256&q=75 HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:06] "GET /static/chunks/webpack-d91acff6684edb66.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:06] "GET /static/chunks/fd9d1056-009d2d77d7923481.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:06] "GET /static/chunks/69-fcfab5d03f85bc33.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:06] "GET /static/chunks/main-app-802dac5169098e0b.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:07] "GET /static/chunks/987-8d0aca4991a01090.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:07] "GET /static/chunks/747-8cf0b916c915bbb5.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:07] "GET /static/chunks/423-0d590daff5094182.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:07] "GET /static/chunks/app/page-3217292ee015c91a.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:07] "GET /static/chunks/723-a2dc065f9de2807e.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:07] "GET /static/chunks/app/layout-a04ea41aed4a098b.js HTTP/1.1" 200 -
127.0.0.1 - - [29/May/2024 00:38:07] "GET /static/css/4c8d1ee479f99e2d.css HTTP/1.1" 304 -
127.0.0.1 - - [29/May/2024 00:38:08] "GET /static/css/c196650da8e3275d.css HTTP/1.1" 304 -
[2024-05-29 00:38:08,407] INFO in server: /search/info: received a request
[2024-05-29 00:38:08,407] INFO in server: /search: finished scrapping and return response
127.0.0.1 - - [29/May/2024 00:38:08] "GET /search/info HTTP/1.1" 200 -
[2024-05-29 00:38:08,408] INFO in server: /data: received a request
[2024-05-29 00:38:08,415] INFO in server: /data: return 589 files response
127.0.0.1 - - [29/May/2024 00:38:08] "GET /data HTTP/1.1" 200 -
```

2. (Optional) Close the webpage on your Browser by clicking the (X) on the tab of the page.



Restart application

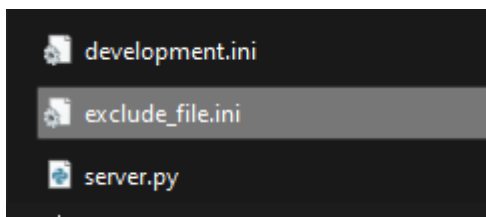
1. Follow step 1 of [Exit application](#) to terminate the program
2. From the application folder, double-click the server.exe file to start the application again. A window will open.
3. After 3-5 minutes, the application screen will show. Indicating the application is ready to access.
4. Reload the webpage on your Browser (or type the link **localhost:8000** if the page is already closed)
5. The restart process is finished and ready to use

Note: You can restart the application also if you find any search process is waiting too long or non-responsive

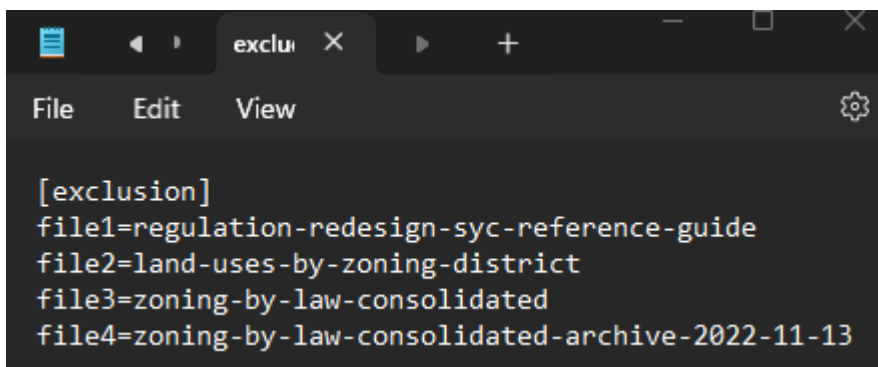
Exclude Downloaded PDF Files

You can exclude any files to prevent them from showing in search result

1. From the application folder, double click exclude_file.ini (If asking what application to open, select 'notepad')



2. Notepad is opened. Certain excluded files are showing in the editor



3. To exclude more files, add the next PDF file name without extension in the format in "file{next number in sequence}={file name}"

For example, if we want to exclude file called policy-plan-broadway.pdf, we should add the text:
file5=policy-plan-broadway in the next line

```
[exclusion]
file1=regulation-redesign-syc-reference-guide
file2=land-uses-by-zoning-district
file3=zoning-by-law-consolidated
file4=zoning-by-law-consolidated-archive-2022-11-13
file5=policy-plan-broadway
```

4. Save the file and close Notepad
5. follow the steps in [Restart the application](#) to restart the application again
6. The file will not show in any search box

Note: To verify the excluded files, you can go to the application terminal (The black screen Window) and find the following text starting with "Excluded files:)

```
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI
* Running on http://127.0.0.1:8000
Press CTRL+C to quit
* Restarting with stat
4 of files is excluded from the list
Excluded files: ['regulation-redesign-syc-reference-guide.pdf', 'land-uses-by-zoning-district.pdf', 'z
* Debugger is active!
* Debugger PIN: 129-119-789
```