

Documentation: 2 Step - Company Status/Active Check

1. Introduction

This script checks the **status of companies** using the **Belgian KBO Public Search website**. It takes a list of company names, searches each one on the website, and retrieves their current status (e.g., active, inactive, not found).

Purpose:

- Automatically check company statuses.
- Save results in an easy-to-read CSV file.

Who is this for?

- Business analysts
 - Researchers
 - Anyone needing bulk company status verification
-

2. What You Need

Software Requirements

- **Python 3.x**
- **Google Chrome**
- **ChromeDriver** (compatible with your Chrome version)

Python Libraries

Make sure these are installed:

```
pip install selenium pandas numpy
```

3. How It Works

1. **Read Data:** Load company names from a CSV file.
2. **Clean Names:** Remove unnecessary words (e.g., "BV," "NV," "SPRL").
3. **Search Website:** Use Selenium to search each company on the KBO website.
4. **Get Status:** Check if the company is active, inactive, or not found.
5. **Save Results:** Save the status and timestamp in a CSV file.

4. Main Parts of the Script

Cleaning Company Names




Purpose: Remove unnecessary words from company names for better search results.

Example:

- **Before:** "ABC BVBA"
- **After:** "ABC"

Searching the Website

For each company:

1. Go to the **KBO Public Search website**.
2. Enter the **clean company name**.
3. Check the results:
 -  **Active:** Company is found and active.
 -  **No Result:** Company is not listed.
 -  **Error:** Something went wrong during the search.

Parallel Processing

The script splits the list of companies into smaller groups and processes them simultaneously. This makes the script faster.

5. How to Run the Script

1. **Prepare the CSV File:**
 - File name: `random_sample_2500_5.csv`
 - Column name: `Name`

Example content:

Name

ABC BV

- XYZ NV
2. **Run the Script:**

Save the script as `company_status.py` and run:

```
python company_status.py
```
 3. **Check the Results:**

After completion, check the output file:

`company_status_random_sample_2500_5.csv`

6. Output File Details

The output CSV will have these columns:

Column Name	Description
OriginalCompanyName	The original company name.
CleanedCompanyName	The cleaned company name.
Status	Status from the KBO website.
Timestamp	Time when status was checked.

Example Output:

```
OriginalCompanyName,CleanedCompanyName,Status,Timestamp  
ABC BV,ABC,Active,2024-09-29 10:15:32  
XYZ NV,XYZ,No result found,2024-09-29 10:18:45
```

7. Error Handling

If something goes wrong while processing a company, it will be marked as:
Status: error

- Timestamp: N/A
 - **Common reasons for errors:**
 - Website timeout
 - Network issues
 - Incorrect company name
-

8. Performance Tips

- **Adjust `num_workers`:** Controls parallel processing. More workers = faster results.
 - **Adjust `sleep_time`:** Time between actions. Reduce cautiously to avoid website blocks.
-

9. Improvements You Can Add

- **Retry Failed Searches:** Try again if a search fails.
- **Better Logging:** Use a logging library for detailed error tracking.

- **Proxy Support:** Use proxies to prevent IP bans on large searches.
-

10. Final Notes

- Make sure ChromeDriver is correctly installed.
 - Run the script on a stable internet connection.
 - Always respect website usage policies.
-

Author: Diāna Koržaviha

Date: 19/11/2024

End of Documentation