**Booking.com Scraper - Projeto de Extracção de Reviews e Informação Gerla de Hoteis**

**Booking Scraper - Guia de Instalação e Utilização**

1. Pré-requisitos

Antes de executares os scripts, certifica-te que tens instalado:

- Python 3.10 ou superior ( na pasta Instalar) – reinicia a máquina

- Google Chrome atualizado

- ChromeDriver compatível com a versão do teu Chrome ( já está na pasta do software)

Este projeto permite-te extrair comentários (reviews) e informação geral de hoteis de vários hotéis listados num ficheiro CSV. Suporta:

* Scroll paginado ("ver mais páginas de comentários")
* Deambular entre páginas de Hoteis
* Aceitação de cookies automaticamente
* Guardar progresso e retomar a recolha após paragem manual (CTRL+C)
* Continuidade sem duplicar comentários
* Guarda todas as Informações Gerais e de Review

**Estrutura do Projeto**

* hoteis.csv — Lista de hotéis (nome e link, link reviews)
* scraper\_reviews\_pagination\_fixed\_v4.py — Script principal para extrair os comentários
* scraper\_hotel\_info.py: Script para extrair informações gerais dos hotéis.
* comentarios\_final\_temp.xlsx — Ficheiro onde os comentários são guardados temporariamente scraper\_reviews\_pagination\_fixed\_v4.py.
* comentarios\_final.xlsx: Ficheiro final consolidado após terminar o scraping scraper\_reviews\_pagination\_fixed\_v4.py.
* hotel\_info.xlsx: Ficheiro final consolidado após terminar o scraping scraper\_hotel\_info.py.
* progress.json — Guarda automaticamente o hotel, página e comentário onde paraste

**Requisitos de Instalação**

1. **Instalar Python 3.10+**
2. **Instalar as bibliotecas obrigatórias**

Uma imagem com texto, captura de ecrã, software, número

Os conteúdos gerados por IA poderão estar incorretos. Uma imagem com texto, captura de ecrã, software, Tipo de letra

Os conteúdos gerados por IA poderão estar incorretos.

**pip install pandas openpyxl selenium beautifulsoup4 webdriver-manager**

**Notas:**

* É usado o ChromeDriver, que é gerido automaticamente pelo webdriver-manager. Não precisas instalar o driver manualmente.

**Como Executar os Scrapers**

1. Certifica-te que tens o hoteis.csv pronto e atualizado.
2. Corre o script no terminal ou no VSCode (ou outro IDE):

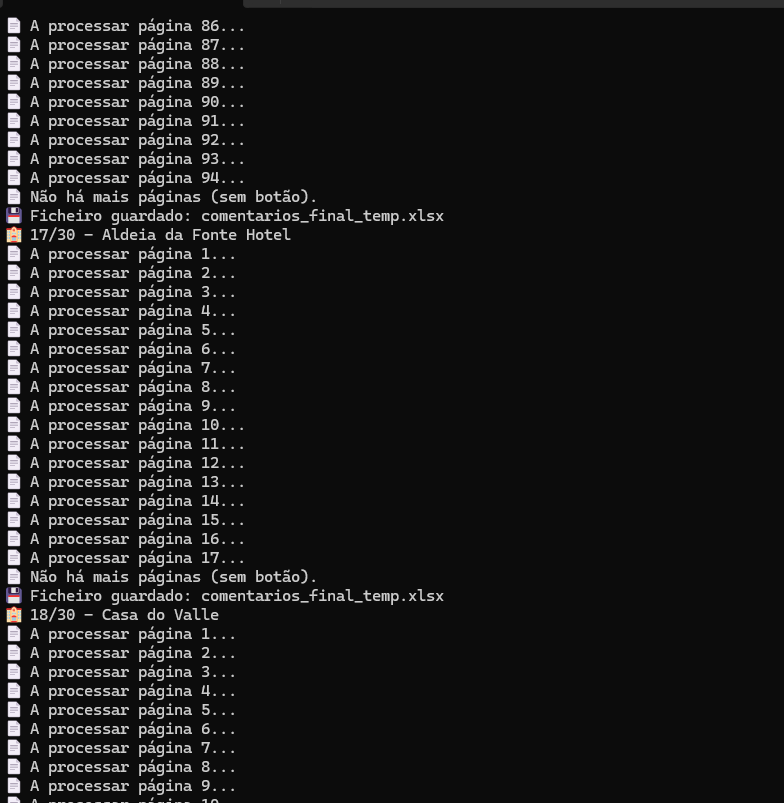
Uma imagem com texto, captura de ecrã, Tipo de letra, número

Os conteúdos gerados por IA poderão estar incorretos.

**Corre o script que pretendes:**  
 - Para comentários: python scraper\_reviews\_pagination\_fixed\_v4.py

**python scraper\_reviews\_pagination\_fixed\_v4.py**  
 - Para informações gerais: python scraper\_hotel\_info.py

**python python scraper\_hotel\_info.py**

****

1. **Aceitar Cookies: O sistema tenta aceitar os cookies automaticamente, mas se aparecer algum pop-up que bloqueie, aceita manualmente na primeira execução.**
2. **O progresso será guardado automaticamente.**
3. **Parar manualmente (opcional):**
   * Se quiseres parar a extração (CTRL+C), o scraper:
     + Salva automaticamente os comentários recolhidos em comentarios\_final\_temp.xlsx
     + Atualiza o progress.json com o último hotel, página e comentário
4. **Retomar depois:**
   * Ao voltares a correr o script, ele detecta o progress.json e continua exatamente do ponto onde paraste, sem duplicar comentários.

**Formato do Ficheiro hoteis.csv**

O ficheiro hoteis.csv tem obrigatoriamente:

|  |  |  |
| --- | --- | --- |
| **Nome** | **Link** | **Link Reviews** |
| Nome do Hotel 1 | Link para o hotel | Link para as reviews |
| Nome do Hotel 2 | Link para o hotel | Link para as reviews |

*Exemplo:*

|  |
| --- |
| **nome,link,link\_reviews** |
| ASPA Studios - Urban Conscious Living,https://www.booking.com/hotel/pt/aspa-sao-paulo.pt-pt.html,https://www.booking.com/hotel/pt/aspa-sao-paulo.pt-pt.html#tab-reviews |
| Hotel Alto Lido,https://www.booking.com/hotel/pt/alto-lido.pt-pt.html,https://www.booking.com/hotel/pt/alto-lido.pt-pt.html#tab-reviews |

...

**Importante:**

* O link deve ir diretamente para a secção de "Reviews" (#tab-reviews).
* O ficheiro 'hoteis.csv' deve manter sempre o mesmo nome.
* Não alteres a estrutura dos ficheiros temporários enquanto o scraping estiver a decorrer.
* Garante que não fechas o Chrome manualmente durante a execução.
* O script gere automaticamente cookies e tentativas de carregamento de páginas.

**Dicas Úteis**

* Certifica-te que tens o Google Chrome atualizado para evitar erros com o driver.
* Se o Booking mudar muito o site no futuro, pode ser necessário ajustar os seletores CSS.
* Guarda cópias regulares do comentarios\_final\_temp.xlsx por segurança.
* Caso haja problemas de clique em botões ou aceitação de cookies, deixa o script clicar automaticamente ou intervém manualmente.
* Se ocorrerem erros de ligação, recomeça o script. O progresso será respeitado.
* Se precisares de atualizar o ChromeDriver, podes usar sites como <https://chromedriver.chromium.org/downloads>.
* **Se quiseres particionar e usar vários hoteis.csv apaga os ficheiros progresso.json e comentarios\_final\_temp.csv assim inicia sempre do principio.**

**Troubleshooting**

* **Erro ao carregar CSV?**
  + Verifica se o hoteis.csv está bem formatado (3 colunas, separadas por vírgula).
* **Erro de cookies?**
  + Aceita manualmente na primeira execução e volta a correr o script.
* **Erro de driver?**
  + Atualiza webdriver-manager com:

***pip install -U webdriver-manager***

Feito com ❤️ por Diogo Amorim