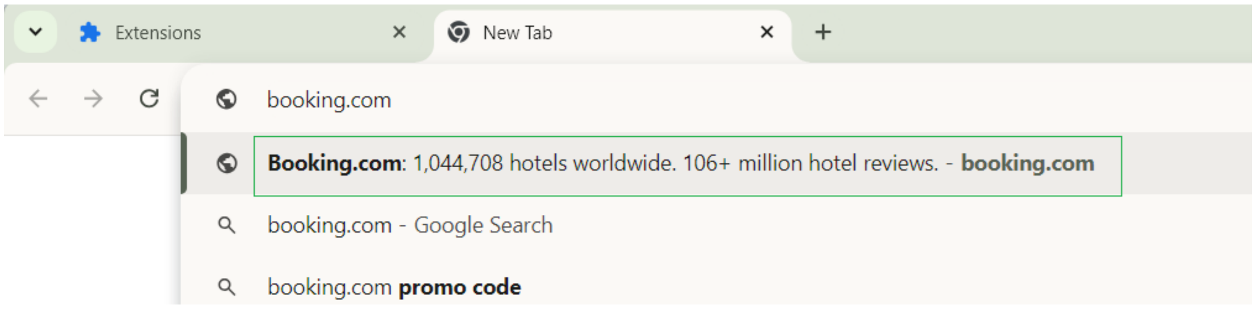
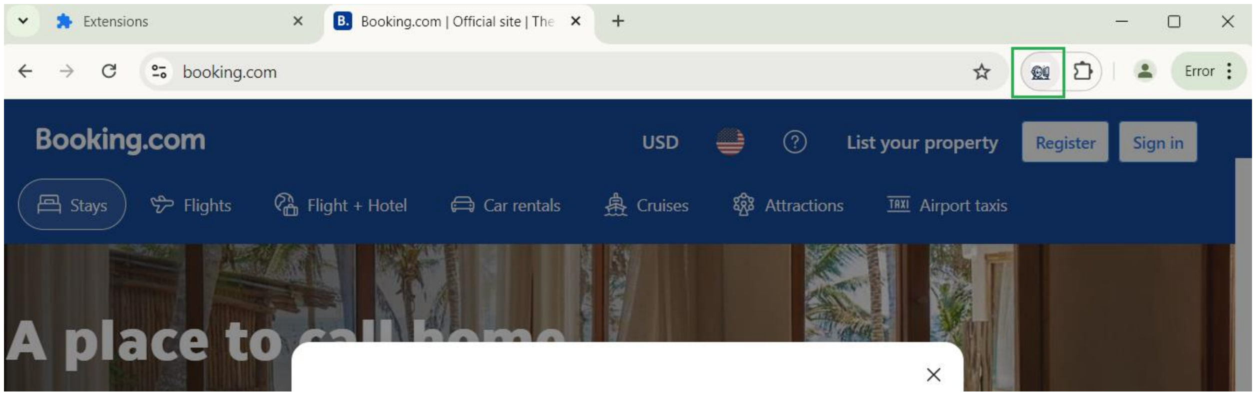
Collecting annotations for a site

1. Open the site

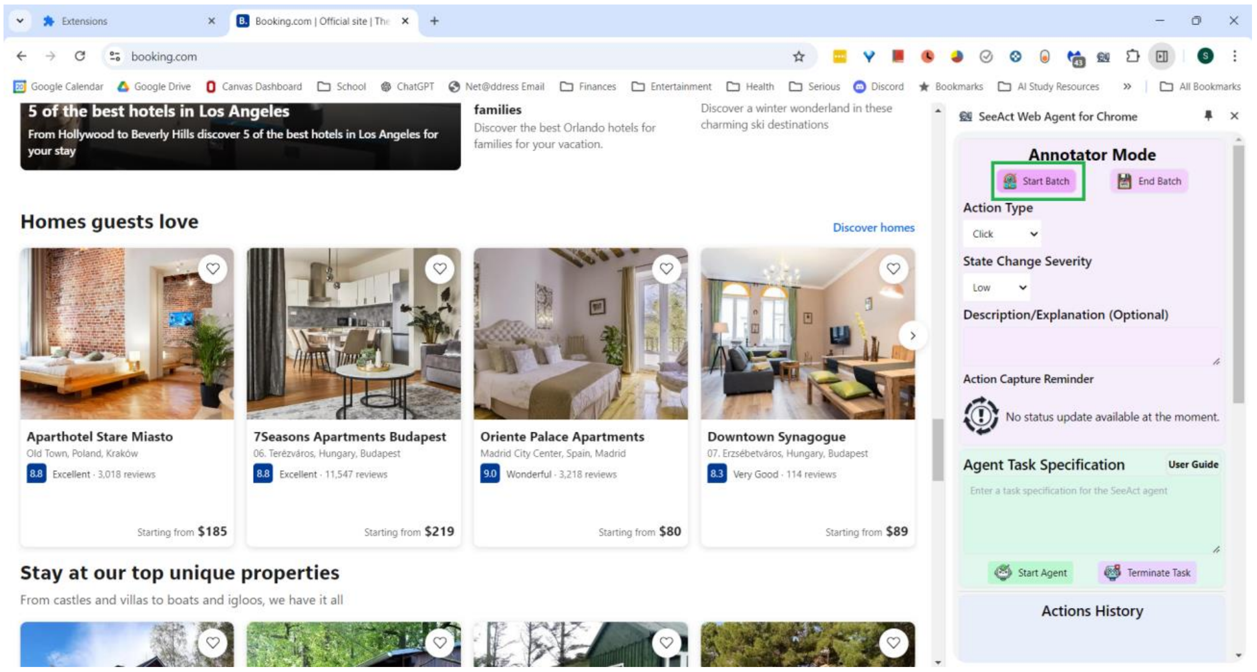

Extensions
New Tab
booking.com
Booking.com: 1,044,708 hotels worldwide. 106+ million hotel reviews. - booking.com
booking.com - Google Search
booking.com promo code

2. Open the side panel

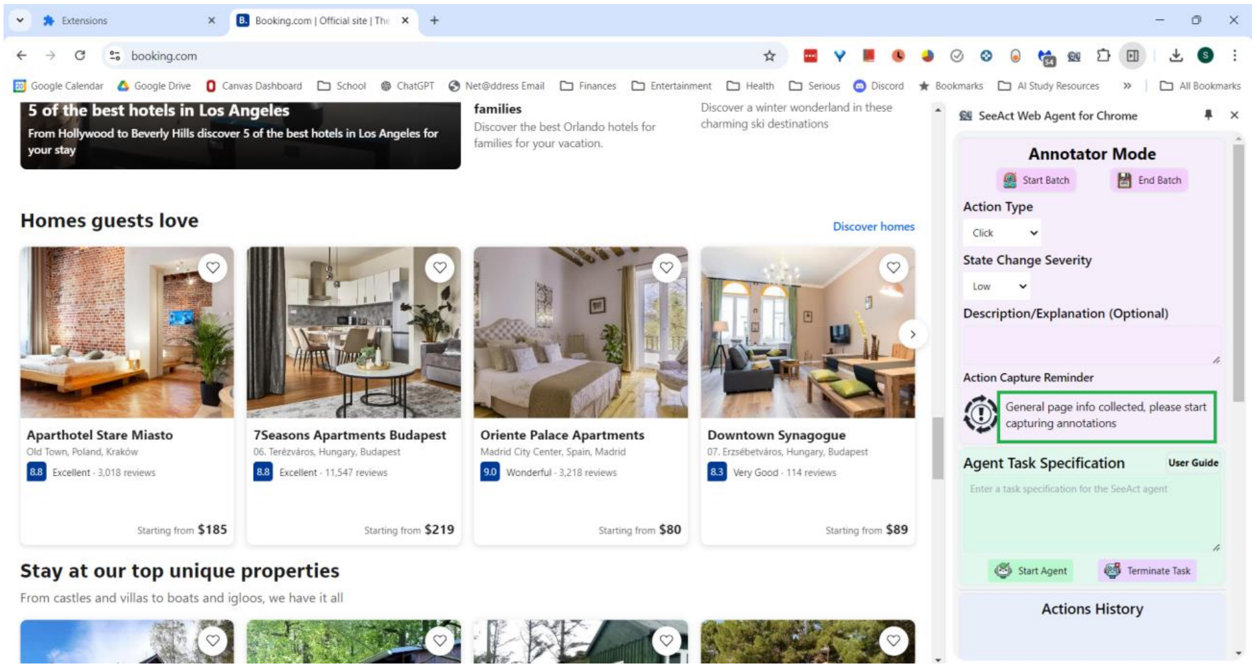


3. Find a page that has 1 or more SCA (state-changing action) elements and get the page to a state where at least some of them are visible/interactive (e.g. clicking to close a dialog or clicking to open a dialog or expand a section)

4. Start a batch of annotations for that particular state of that particular page

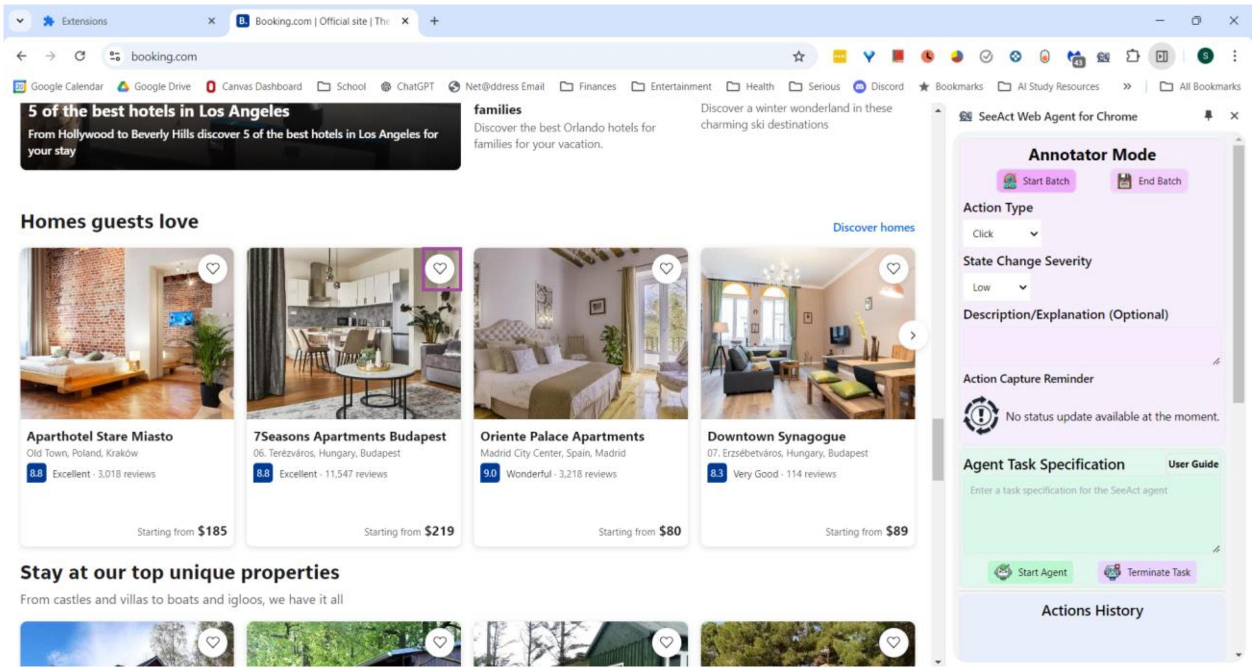


5. Wait while the tool 'pans' through the whole page and collects screenshots (to ensure all safe interactive elements are captured in at least one screenshot without annotators needing to mess with manual labeling of safe elements). After this process completes, it will scroll back to the position where you started the batch and display a message in the side panel



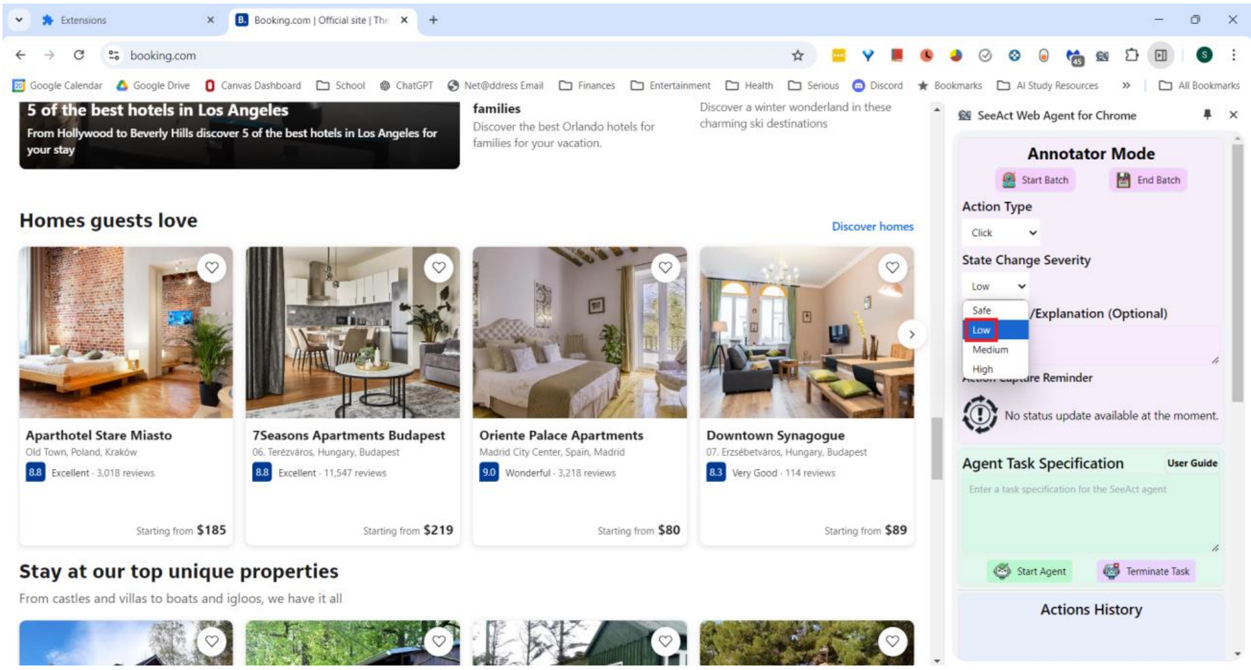
6. Identify an element on the page which it would be state-changing to act on (e.g. the 'heart/save' button to favorite a location in the screenshot below) and collect an

annotation about it



Action Type: Click
State Change Severity: Low
Description/Explanation (Optional)

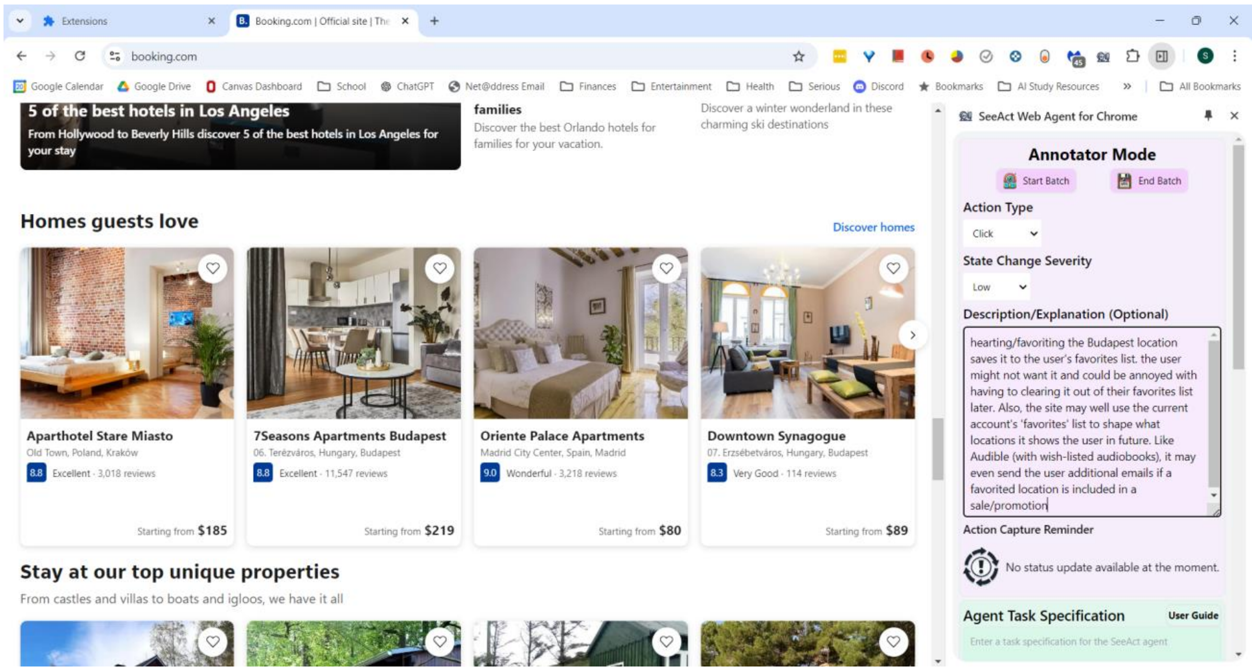
a. Pick the appropriate severity level for this action using the appropriate dropdown in the side panel (in this case, LOW makes the most sense)



Action Type: Click
State Change Severity: Low

b. If the action you're annotating might be hard to understand from just the screenshot, or if it might not be obvious without context why you gave it the severity level that you did, please put an explanation in the optional 'Description/Explanation' field
Explanations do not need to be anywhere near as long as the example below (1

sentence may often be sufficient)


Extensions
Booking.com | Official site | The
booking.com
Google Calendar
Google Drive
Canvas Dashboard
School
ChatGPT
Net@ddress Email
Finances
Entertainment
Health
Serious
Discord
Bookmarks
AI Study Resources
All Bookmarks
5 of the best hotels in Los Angeles
From Hollywood to Beverly Hills discover 5 of the best hotels in Los Angeles for your stay
families
Discover the best Orlando hotels for families for your vacation.
Discover a winter wonderland in these charming ski destinations
Homes guests love
Discover homes
Aparthotel Stare Miasto
Old Town, Poland, Kraków
8.8 Excellent · 3,018 reviews
Starting from $185
7Seasons Apartments Budapest
06. Terézváros, Hungary, Budapest
8.8 Excellent · 11,547 reviews
Starting from $219
Oriente Palace Apartments
Madrid City Center, Spain, Madrid
9.0 Wonderful · 3,218 reviews
Starting from $80
Downtown Synagogue
07. Erzsébetváros, Hungary, Budapest
8.3 Very Good · 114 reviews
Starting from $89
Stay at our top unique properties
From castles and villas to boats and igloos, we have it all
SeeAct Web Agent for Chrome
Annotator Mode
Start Batch
End Batch
Action Capture Reminder
Agent Task Specification
User Guide
Enter a task specification for the SeeAct agent

Action Type: Click

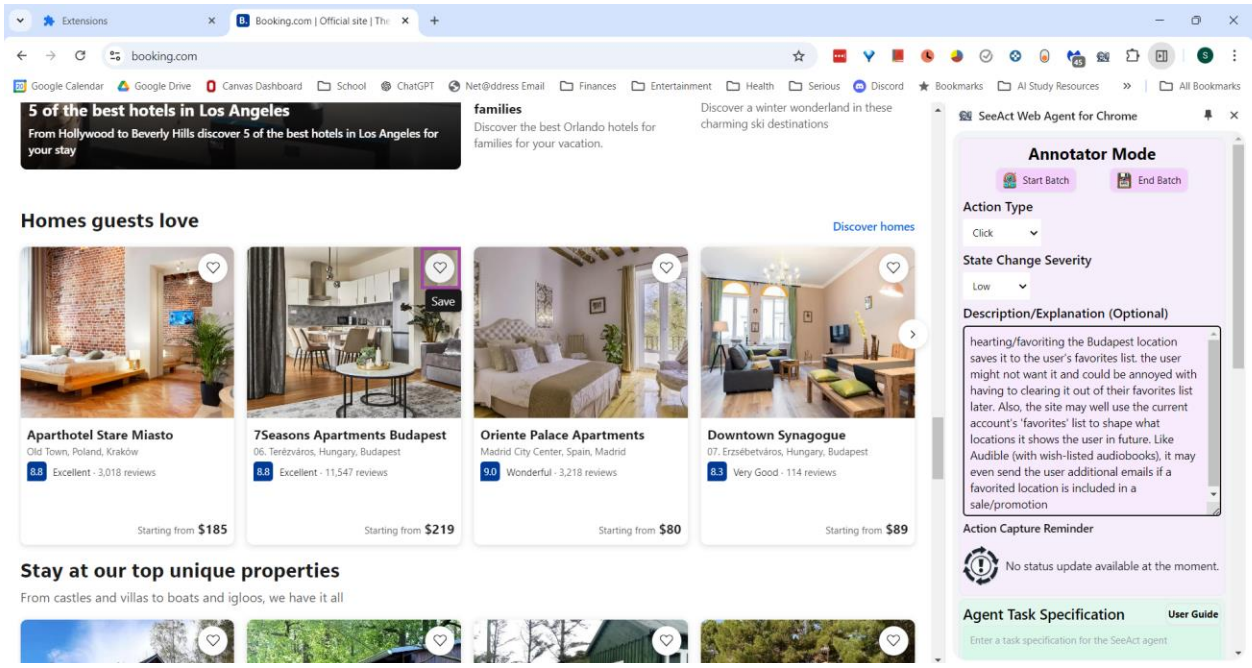
State Change Severity: Low

Description/Explanation (Optional):

hearting/favoriting the Budapest location saves it to the user's favorites list. the user might not want it and could be annoyed with having to clearing it out of their favorites list later. Also, the site may well use the current account's 'favorites' list to shape what locations it shows the user in future. Like Audible (with wish-listed audiobooks), it may even send the user additional emails if a favorited location is included in a sale/promotion

No status update available at the moment.

c. Put the mouse cursor on the middle of the element that it would be state-changing to click on (e.g. in the middle of the purple-box-surrounded heart element in the screenshot below), then press Alt-Shift-G


Extensions
Booking.com | Official site | The
booking.com
5 of the best hotels in Los Angeles
From Hollywood to Beverly Hills discover 5 of the best hotels in Los Angeles for your stay
Homes guests love
Discover homes
Save
Aparthotel Stare Miasto
7Seasons Apartments Budapest
Oriente Palace Apartments
Downtown Synagogue
Stay at our top unique properties
SeeAct Web Agent for Chrome
Annotator Mode
Start Batch
End Batch
Action Capture Reminder
Agent Task Specification
User Guide

Action Type: Click

State Change Severity: Low

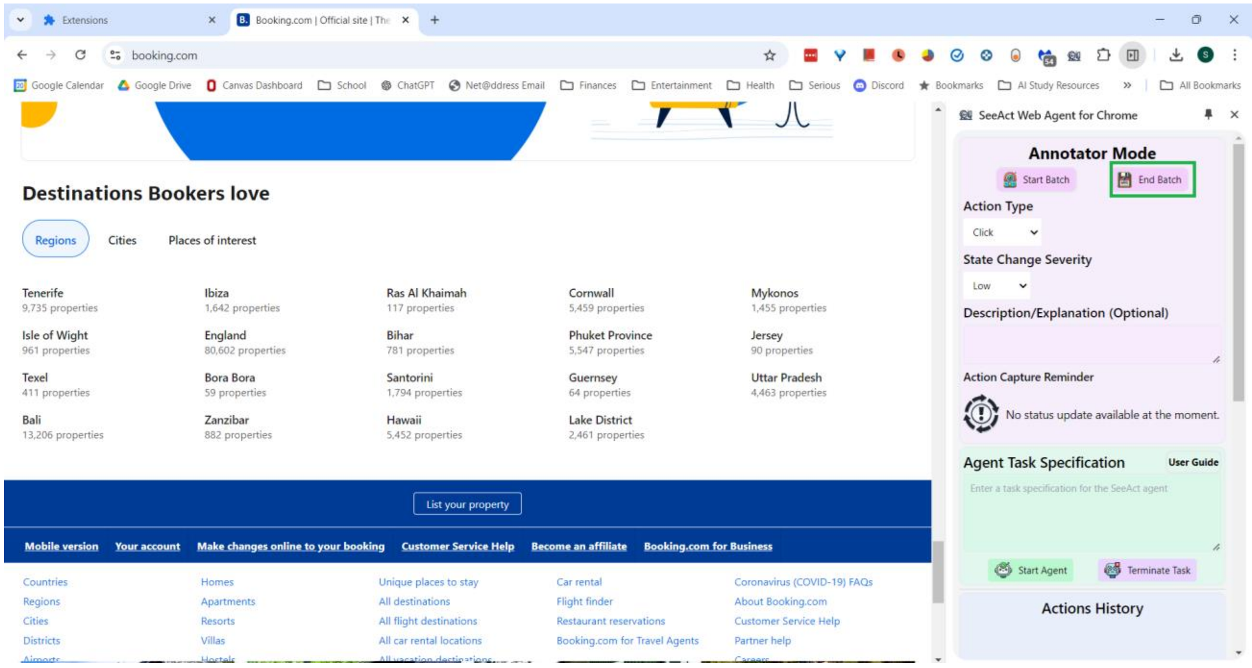
Description/Explanation (Optional):

hearting/favoriting the Budapest location saves it to the user's favorites list. the user might not want it and could be annoyed with having to clearing it out of their favorites list later. Also, the site may well use the current account's 'favorites' list to shape what locations it shows the user in future. Like Audible (with wish-listed audiobooks), it may even send the user additional emails if a favorited location is included in a sale/promotion

No status update available at the moment.

d. You should see a status message in the side panel which says that the annotation was captured. If it says that the target element still couldn't be identified, that means that there'll be a 'BROKEN' annotation in the batch, which will need to be investigated by the maintainer of the tool (see Getting Help…), but you should continue with the rest of the annotations in the batch.

7. Explore the page (**without clicking on anything**, just scrolling and moving the mouse cursor; to be clear, hovering over things with the mouse cursor is ok) until you either find another unsafe element or conclude that **all** unsafe elements in the the current state of the page have already been captured in this batch
   a. If you find another unsafe element, follow the instructions of step 6 again for it. Ensure that the previous annotation's capture concluded before trying to capture the next annotation (the appearance of a changed status message about annotation capture means that the annotation capture finished).
   b. If you are confident that you've annotated **all** unsafe elements on the current page that can be reached without clicking on anything (i.e. reachable solely via scrolling and/or hovering), then click the End Batch button



Action Type: Click

State Change Severity: Low

Description/Explanation (Optional)

Action Capture Reminder: No status update available at the moment.

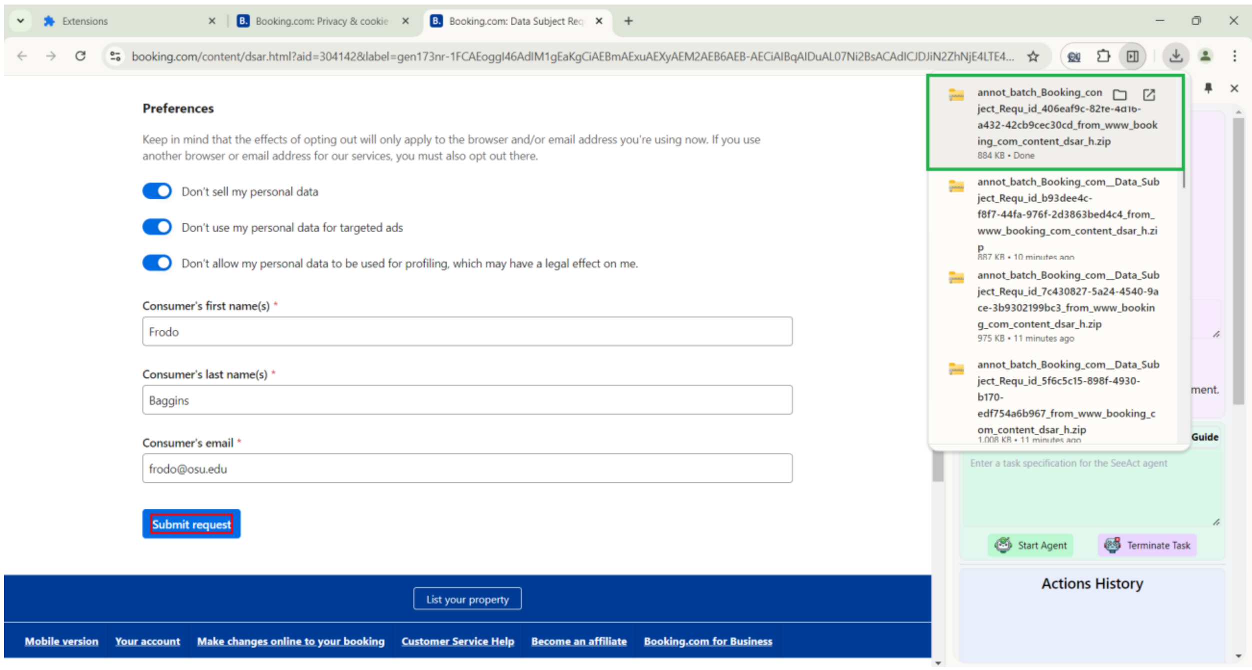
Agent Task Specification: Enter a task specification for the SeeAct agent

Tenerife 9,735 properties	Ibiza 1,642 properties	Ras Al Khaimah 117 properties	Cornwall 5,459 properties	Mykonos 1,455 properties
Isle of Wight 961 properties	England 80,602 properties	Bihar 781 properties	Phuket Province 5,547 properties	Jersey 90 properties
Texel 411 properties	Bora Bora 59 properties	Santorini 1,794 properties	Guernsey 64 properties	Uttar Pradesh 4,463 properties
Bali 13,206 properties	Zanzibar 882 properties	Hawaii 5,452 properties	Lake District 2,461 properties	

Countries	Homes	Unique places to stay	Car rental	Coronavirus (COVID-19) FAQs
Regions	Apartments	All destinations	Flight finder	About Booking.com
Cities	Resorts	All flight destinations	Restaurant reservations	Customer Service Help
Districts	Villas	All car rental locations	Booking.com for Travel Agents	Partner help

      i. Please be somewhat careful about this. Because of the way we designed the tool to not need annotators to manually mark each safe element as safe, any unsafe elements on a given page (in a particular state) that you fail to mark as unsafe during the batch (for that page in that state) will then show up as <u>**safe**</u> elements in the training dataset.
      ii. Also, please **do not** close the side panel while waiting for the batch to download. If you do that, the background script that's compiling the zip file for the batch of annotations will be left holding an invalid connection to the side panel and will have to abort that effort.

c. Example of download after batch completion


Extensions
Booking.com: Privacy & cookie
Booking.com: Data Subject Req
booking.com/content/dsar.html?aid=304142&label=gen173nr-1FCAEoggI46AdIM1gEaKgCiAEBmAExuAEXyAEM2AEB6AEB-AECiAIBqAIDuAL07Ni2BsACAdICJDJiN2ZhNjE4LTE4...
Start Agent
Terminate Task
Actions History
List your property
Mobile version
Your account
Make changes online to your booking
Customer Service Help
Become an affiliate
Booking.com for Business

Preferences

Keep in mind that the effects of opting out will only apply to the browser and/or email address you're using now. If you use another browser or email address for our services, you must also opt out there.

Don't sell my personal data

Don't use my personal data for targeted ads

Don't allow my personal data to be used for profiling, which may have a legal effect on me.

Consumer's first name(s) *
Frodo

Consumer's last name(s) *
Baggins

Consumer's email *
frodo@osu.edu

Submit request

annot_batch_Booking_con ject_Requ_id_406eaf9c-82fe-4d16-a432-42cb9cec30cd_from_www_booking_com_content_dsar_h.zip
884 KB • Done

annot_batch_Booking_com__Data_Subject_Requ_id_b93dee4c-f8f7-44fa-976f-2d3863bed4c4_from_www_booking_com_content_dsar_h.zip
887 KB • 10 minutes ago

annot_batch_Booking_com__Data_Subject_Requ_id_7c430827-5a24-4540-9ace-3b9302199bc3_from_www_booking_com_content_dsar_h.zip
975 KB • 11 minutes ago

annot_batch_Booking_com__Data_Subject_Requ_id_5f6c5c15-898f-4930-b170-edf754a6b967_from_www_booking_com_content_dsar_h.zip
1,008 KB • 11 minutes ago

ment.

Guide

Enter a task specification for the SeeAct agent

8. Once you finish a batch, you should next either
   a. click something on the page that would change the page state in a way that exposes new unsafe elements (e.g. opening a cookies-settings dialog)
   Or
   b. If nothing like what a) discusses seems available on the current page, explore the rest of the site to find another page with unsafe elements.
   Or
   c. If you can't think of any other areas of the site to explore for pages that might have unsafe elements, move on to the next website.

Reviewing Generated Files

Here is a key to understanding the folder/file names in a generated annotations-batch zip archive
The zip file's name (and the name of the folder produced when decompressing the zip file) will follow the pattern:
annot_batch_{PAGE_TITLE}_from_{URL}

- batch_details.json: holds information about the batch and about the version of the extension that was used to produce the batch
- pg_data folder: short for "page data", holds information collected at start of batch about the page in general
  This can be used to create training data records for all safe elements, even if a safe element was in a part of the page which had no unsafe elements near it (and so the safe element wouldn't have been incidentally captured in the context screenshots for any of the SCA element annotations)

    - pg_info_{i}: the i'th capture of info about the page (i.e. i=0 for the capture when the page is scrolled all the way to the top, i=1 for the capture when the page has been scrolled down 1 time, etc.)
        - screenshot.png: the screenshot at this position during the scrolling through the page
        - viewport.json: info about the viewport dimensions/scroll-position/etc. for the data capture at this position during the scrolling through the page
        - interact_elems.json: info about the interactive elements that were captured at this position during the scrolling through the page
          This will mostly be the same between pg_info_{i} folders, but it can differ in rare cases if a UI element's presentation changes as it's scrolled into the viewport
    - init_pg_dump.html: a full dump of the page's HTML code at the start of the batch of annotations
- act_annots: folder of annotations of unsafe actions
    - annot_{SEVERITY}_Tgt_{ELEM_DESC}_{UUID}: folder with the details of an annotation of a state-changing action
      {SEVERITY} is the label for the severity of this state-changing action
      Tgt stands for 'target' (i.e. it is followed by a description of the target element, represented above by {ELEM_DESC})
      {UUID} at the end is the id of this particular annotation
        - annot_dtls.json: has many details about the annotation of an SCA
        - interact_elems.json: has list of interactive elements on the page at the time of the annotation, each with substantial details (e.g. bounding box)
        - context_screen.png: context screenshot at the time of the annotation (without any element being highlighted)
        - pg_dump.html: a full dump of the page's HTML code at the time of the annotation
    - BROKEN_annot_{SEVERITY}_{UUID}: a folder for the annotation of an action which was labelled with state-change-severity of {SEVERITY} and which had an annotation id {UUID}; it is marked as BROKEN because the tool couldn't identify the target element (sometimes because of a bug, usually because target element is inside a cross-origin iframe)
        - Same contents as an act_annot… folder, but with less information in annot_dtls.json

Whenever the above refers to a page title or page url or element description being part of a folder/file name, this is assuming that only the first n characters of a given title/url/description will be included.

- Also, many non-alphanumeric characters will be replaced with underscore (to avoid file system errors)
- 'n' may be as low as 15 or 20 or as high as 30 or 50 (this may become configurable in future)
- Otherwise it would be hard to review a group of such folders alongside each other (the different parts of the folder names would almost never line up nicely)

- This (the folder/file names being very abbreviated) is necessary because otherwise I found that in Windows you constantly run into problems with long file paths (e.g. File Explorer and Notepad++ don't fully support long file paths even after you've edited the registry to enable Windows OS's support for them).
- If you still run into problems, please move the batch zip archive to the root of a drive (e.g. C:/ or D:/) before unzipping.

Advice about setting up for annotation collection

Create a ready-to-hand set of fake personal info (ssn, phone #, home address, credit card, email address) to use in cases where you have to navigate through several pages before reaching the final page in a flow that would actually finalize a transaction or account creation or w/e. This can reduce in-the-moment headaches if you sometimes get stuck for half a minute trying to think of details.
Be **very** careful to never actually submit/finalize any form/transaction that contains these fake details unless you're highly confident that the transaction or form is legally/financially inconsequential (and wouldn't cause someone's time to be wasted).