# Manual: how to use the Expedia spider

## Requirements

python，scrapy，selenium，geckodriver，firefox

## What does it do

Crawl hotel info and room price

Hotel: name, area, zipcode, star, rating, rates, room, room size, price, bed(s), sleep guests, address, url

## How to:

\tu> scrapy crawl quotes -o filename.csv

not including sold out hotel

## Change csv column order

Locate to \tu\tu，find settings.py in 'FEED\_EXPORT\_FIELDS'

## Change target city/date

open tu/spiders/quotes\_spider.py

change the crucial info：

# destination, travel dates, crawler run times.

destInput = "New York"

timeinterval = 1

crawltimes = 30

# timeinterval is travel dates, = 1 means you stay for 1 night, and so on.

# crawltimes is how many time search result will be crawled.

# after one date period results finished, crawl (next day + dates) result.

other properties:

# Tip: set "345" to crawl 3/4/5 class, or just one "4" for 4 class only

# for some unknown reason, expedia may not show 1 and 2 star filter for clickble

# besides, if you choose 3 then 5, 4 will also be choosed and non-cancellable

# set to 0 to not use any filter.

hotelstar = "0"

# set 1 to choose 1 adult for search, set 0 to use default 2 adults.

oneP = 0

# if it is a sponsored result, set 1 to save as usual(may cause duplicate later), set 0 to drop this result.

# the sponsored hotel may show up as the first 2 and last 2 result.

sponsorAval = 0

# if zipcode info is unknown, set 1 to save it as 'null', set 0 to drop this hotel

hasZip = 1

# if size info is unknown, set 1 to save it as 'null', set 0 to drop this room info row

# about 10%-35%(vary) result items is null, blame expedia

sizeUnknow = 0

# size can be 0 in some very rare condition, set 1 to save as 0, set 0 to drop this room info row

sizeBezero = 0

# if size like"275-350", set 0 to save as 275, set 1 to save as 312, set 2 save as "275-350"

# Note: items like this less than 2%

sizeAvg = 0

# if rating info is unknown, set 1 to save it as 'null', set 0 to drop this hotel

# Note: no rating is a very rare condition

ratingUnknow = 0

# if bed info is unknown, set 1 to save it as 'null', set 0 to drop this room info row

# Note: no bed info is a very rare condition

bedUnknow = 1

# if guests info is unknown, set 1 to save it as 'null', set 0 to drop this room info row

# Note: no guests info is a very rare condition

guestsUnknow = 1