

Scraping craigslist

Without being banned for life?...

Robert D. West

scraping craigslist

Setup:

Run the following from your favorite directory to begin using `automated_craigslist_search`

```
$ git clone git@github.com:robertdavidwest/api-automated_craigslist_search.git #assuming ssh install
$ cd automated_craigslist_search
$ python setup.py install
```

Packages needed: urllib2, bs4, pandas, email

What does it do?

Automates craigslist searches, stores the results in a `pandas.DataFrame`, allows you to e-mail results as html

scraping craigslist

Setup:

Run the following from your favorite directory to begin using `automated_craigslist_search`

```
$ git clone git@github.com:robertdavidwest/api-automated_craigslist_search.git #assuming ssh install
$ cd automated_craigslist_search
$ python setup.py install
```

Packages needed: urllib2, bs4, pandas, email

What does it do?

Automates craigslist searches, stores the results in a `pandas.DataFrame`, allows you to e-mail results as html

scraping craigslist

Example:

```
In [28]: import automated_craigslist_search.connect_to_craigslist as connect_to_craigslist
```

```
In [29]: df = connect_to_craigslist.search_craigslist('katy perry',min_value =  
100,max_value=200,category='tickets',city='newyork')
```

```
In [30]: print df.head()
```

	Date	Location	Price	\
0	Jun 26	(Midtown East)	\$125	
1	Jun 26	(Hoboken, NJ)	\$115	
2	Jun 26	(Midtown West)	\$200	
3	Jun 25	(Upper East Side)	\$150	
4	Jun 25	(Upper East Side)	\$150	

	Results	urls
0	Katy Perry July 9th MSG	/mnh/tix/4534246055.html
1	KATY PERRY/7-11/PRUDENTIAL CENTER	/jsy/tix/4503519503.html
2	The Prismatic World Tour (Katy Perry) M.S.G	/mnh/tix/4518250309.html
3	Katy Perry at MSG 7/9	/mnh/tix/4538330862.html
4	Katy Perry at MSG 7/9	/mnh/tix/4538175626.html

scraping craigslist

- Activity Partners

scraping craigslist

- Activity Partners
- Across 7 major US Cities: Chicago, Houston, Los Angeles, New York, Philadelphia, Phoenix, San Antonio

scraping craigslist

- Activity Partners
- Across 7 major US Cities: Chicago, Houston, Los Angeles, New York, Philadelphia, Phoenix, San Antonio
- Found the top 20 most commonly occurring search words (limit of 1 word per search).

scraping craigslist

- Activity Partners
- Across 7 major US Cities: Chicago, Houston, Los Angeles, New York, Philadelphia, Phoenix, San Antonio
- Found the top 20 most commonly occurring search words (limit of 1 word per search).
- Started with a `pandas.DataFrame`

scraping craigslist

- Activity Partners
- Across 7 major US Cities: Chicago, Houston, Los Angeles, New York, Philadelphia, Phoenix, San Antonio
- Found the top 20 most commonly occurring search words (limit of 1 word per search).
- Started with a `pandas.DataFrame`
- Cleaned the data: removed duplicate entries, made lower case, removed punctuation and numbers

scraping craigslist

- Activity Partners
- Across 7 major US Cities: Chicago, Houston, Los Angeles, New York, Philadelphia, Phoenix, San Antonio
- Found the top 20 most commonly occurring search words (limit of 1 word per search).
- Started with a `pandas.DataFrame`
- Cleaned the data: removed duplicate entries, made lower case, removed punctuation and numbers
- Removed common words

scraping craigslist

- Removed common words:

```
In [33]: print std_word_set
set(['all', 'just', 'being', 'over', 'both', 'through', 'yourselves', 'its',
'before', 'herself', 'had', 'should', 'to', 'only', 'under', 'ours', 'has', 'do',
'them', 'his', 'very', 'they', 'not', 'during', 'now', 'him', 'nor', 'did', 'this',
'she', 'each', 'further', 'where', 'few', 'because', 'doing', 'some', 'are', 'our',
'ourselves', 'out', 'what', 'for', 'while', 'does', 'above', 'between', 't', 'be',
'we', 'who', 'were', 'here', 'hers', 'by', 'on', 'about', 'of', 'against', 's', 'or',
'own', 'into', 'yourself', 'down', 'your', 'from', 'her', 'their', 'there', 'been',
'whom', 'too', 'themselves', 'was', 'until', 'more', 'himself', 'that', 'but', 'don',
'with', 'than', 'those', 'he', 'me', 'myself', 'these', 'up', 'will', 'below', 'can',
'theirs', 'my', 'and', 'then', 'is', 'am', 'it', 'an', 'as', 'itself', 'at', 'have',
'in', 'any', 'if', 'again', 'no', 'when', 'same', 'how', 'other', 'which', 'you',
'after', 'most', 'such', 'why', 'a', 'off', 'i', 'yours', 'so', 'the', 'having',
'once'])
```

scraping craigslist

- Removed common words: The rest by trial and error:

```
location_words = ['manhattan', 'midtown', "nyc", "houston", "san", "antonio",  
"pheonix", "city", "chicago", "phoenix"]  
needy_words = ["looking", "seeking", "want", "wanted", "needed", "need", "help"]  
others = ['sick', 'action', "new", 'opening', "starts", "players", "player", "free",  
"partner", "team", "group", "july", "get", "th"]
```

scraping craigslist

	Chicago	Houston	Los Angeles	New York	Philadelphia	Phoenix	San Antonio
1	tennis (25)	workout (12)	tennis (38)	tennis (44)	tennis (25)	softball (36)	league (9)
2	female (10)	join (9)	lessons (15)	poker (30)	softball (15)	soccer (24)	tennis (9)
3	soccer (9)	trainer (8)	female (15)	female (30)	training (11)	coed (17)	basketball (7)
4	lessons (7)	training (8)	softball (14)	game (23)	coed (9)	female (16)	fitness (6)
5	softball (6)	fitness (8)	trainer (14)	lessons (19)	fitness (8)	personal (14)	coed (6)
6	play (5)	tennis (8)	training (13)	english (19)	partners (8)	training (13)	softball (4)
7	coed (5)	soccer (8)	league (13)	exchange (18)	female (8)	league (12)	training (4)
8	partners (5)	personal (8)	coed (12)	park (17)	friends (7)	football (12)	workout (4)
9	league (5)	partners (7)	workout (12)	language (17)	workout (7)	play (11)	football (4)
10	arts (4)	ladies (6)	pick (11)	softball (14)	personal (7)	summer (10)	work (4)
11	martial (4)	buddy (6)	personal (11)	central (14)	boxing (6)	indoor (10)	mens (4)
12	summer (4)	friends (6)	classes (10)	personal (12)	learn (6)	workout (10)	gym (3)
13	buddy (3)	certified (5)	soccer (10)	yoga (12)	play (6)	male (9)	like (3)
14	men (3)	activity (5)	gym (10)	club (10)	fit (6)	night (9)	advice (3)
15	baseball (3)	weight (5)	buddy (10)	best (10)	soccer (5)	tennis (8)	friends (3)
16	fitness (3)	lose (5)	boxing (9)	summer (10)	make (5)	youth (8)	camps (3)
17	workout (3)	study (4)	dance (8)	play (9)	month (5)	kickers (7)	flair (3)
18	adult (3)	coach (4)	fun (8)	korean (9)	bootcamp (5)	flag (7)	events (3)
19	amateur (3)	football (4)	mens (8)	sunday (9)	baseball (5)	leagues (7)	online (3)
20	make (3)	lessons (4)	people (8)	training (9)	running (5)	lil (7)	tournament (3)

scraping craigslist

	Sports						
	Chicago	Houston	Los Angeles	New York	Philadelphia	Phoenix	San Antonio
1	tennis (25)	workout (12)	tennis (38)	tennis (44)	tennis (25)	softball (36)	league (9)
2	female (10)	join (9)	lessons (15)	poker (30)	softball (15)	soccer (24)	tennis (9)
3	soccer (9)	trainer (8)	female (15)	female (30)	training (11)	coed (17)	basketball (7)
4	lessons (7)	training (8)	softball (14)	game (23)	coed (9)	female (16)	fitness (6)
5	softball (6)	fitness (8)	trainer (14)	lessons (19)	fitness (8)	personal (14)	coed (6)
6	play (5)	tennis (8)	training (13)	english (19)	partners (8)	training (13)	softball (4)
7	coed (5)	soccer (8)	league (13)	exchange (18)	female (8)	league (12)	training (4)
8	partners (5)	personal (8)	coed (12)	park (17)	friends (7)	football (12)	workout (4)
9	league (5)	partners (7)	workout (12)	language (17)	workout (7)	play (11)	football (4)
10	arts (4)	ladies (6)	pick (11)	softball (14)	personal (7)	summer (10)	work (4)
11	martial (4)	buddy (6)	personal (11)	central (14)	boxing (6)	indoor (10)	mens (4)
12	summer (4)	friends (6)	classes (10)	personal (12)	learn (6)	workout (10)	gym (3)
13	buddy (3)	certified (5)	soccer (10)	yoga (12)	play (6)	male (9)	like (3)
14	men (3)	activity (5)	gym (10)	club (10)	fit (6)	night (9)	advice (3)
15	baseball (3)	weight (5)	buddy (10)	best (10)	soccer (5)	tennis (8)	friends (3)
16	fitness (3)	lose (5)	boxing (9)	summer (10)	make (5)	youth (8)	camps (3)
17	workout (3)	study (4)	dance (8)	play (9)	month (5)	kickers (7)	flair (3)
18	adult (3)	coach (4)	fun (8)	korean (9)	bootcamp (5)	flag (7)	events (3)
19	amateur (3)	football (4)	mens (8)	sunday (9)	baseball (5)	leagues (7)	online (3)
20	make (3)	lessons (4)	people (8)	training (9)	running (5)	lil (7)	tournament (3)

scraping craigslist

	Average Winter Low						2
	18	23	24	40	43.2	46	48
	Chicago	New York	Philadelphia	San Antonio	Houston	Phoenix	Los Angeles
	tennis (25)	tennis (44)	tennis (25)	league (9)	workout (12)	softball (36)	tennis (38)
	female (10)	poker (30)	softball (15)	tennis (9)	join (9)	soccer (24)	lessons (15)
	soccer (9)	female (30)	training (11)	basketball (7)	trainer (8)	coed (17)	female (15)
	lessons (7)	game (23)	coed (9)	fitness (6)	training (8)	female (16)	softball (14)
	softball (6)	lessons (19)	fitness (8)	coed (6)	fitness (8)	personal (14)	trainer (14)
	play (5)	english (19)	partners (8)	softball (4)	tennis (8)	training (13)	training (13)
	coed (5)	exchange (18)	female (8)	training (4)	soccer (8)	league (12)	league (13)
	partners (5)	park (17)	friends (7)	workout (4)	personal (8)	football (12)	coed (12)
1	league (5)	language (17)	workout (7)	football (4)	partners (7)	play (11)	workout (12)
1	arts (4)	softball (14)	personal (7)	work (4)	ladies (6)	summer (10)	pick (11)
1	martial (4)	central (14)	boxing (6)	mens (4)	buddy (6)	indoor (10)	personal (11)
1	summer (4)	personal (12)	learn (6)	gym (3)	friends (6)	workout (10)	classes (10)
1	buddy (3)	yoga (12)	play (6)	like (3)	certified (5)	male (9)	soccer (10)
1	men (3)	club (10)	fit (6)	advice (3)	activity (5)	night (9)	gym (10)
1	baseball (3)	best (10)	soccer (5)	friends (3)	weight (5)	tennis (8)	buddy (10)
1	fitness (3)	summer (10)	make (5)	camps (3)	lose (5)	youth (8)	boxing (9)
1	workout (3)	play (9)	month (5)	flair (3)	study (4)	kickers (7)	dance (8)
1	adult (3)	korean (9)	bootcamp (5)	events (3)	coach (4)	flag (7)	fun (8)
2	amateur (3)	sunday (9)	baseball (5)	online (3)	football (4)	leagues (7)	mens (8)
	make (3)	training (9)	running (5)	tournament (3)	lessons (4)	lil (7)	people (8)

scraping craigslist

	Sex						
	Chicago	Houston	Los Angeles	New York	Philadelphia	Phoenix	San Antonio
1	tennis (25)	workout (12)	tennis (38)	tennis (44)	tennis (25)	softball (36)	league (9)
2	female (10)	join (9)	lessons (15)	poker (30)	softball (15)	soccer (24)	tennis (9)
3	soccer (9)	trainer (8)	female (15)	female (30)	training (11)	coed (17)	basketball (7)
4	lessons (7)	training (8)	softball (14)	game (23)	coed (9)	female (16)	fitness (6)
5	softball (6)	fitness (8)	trainer (14)	lessons (19)	fitness (8)	personal (14)	coed (6)
6	play (5)	tennis (8)	training (13)	english (19)	partners (8)	training (13)	softball (4)
7	coed (5)	soccer (8)	league (13)	exchange (18)	female (8)	league (12)	training (4)
8	partners (5)	personal (8)	coed (12)	park (17)	friends (7)	football (12)	workout (4)
9	league (5)	partners (7)	workout (12)	language (17)	workout (7)	play (11)	football (4)
10	arts (4)	ladies (6)	pick (11)	softball (14)	personal (7)	summer (10)	work (4)
11	martial (4)	buddy (6)	personal (11)	central (14)	boxing (6)	indoor (10)	mens (4)
12	summer (4)	friends (6)	classes (10)	personal (12)	learn (6)	workout (10)	gym (3)
13	buddy (3)	certified (5)	soccer (10)	yoga (12)	play (6)	male (9)	like (3)
14	men (3)	activity (5)	gym (10)	club (10)	fit (6)	night (9)	advice (3)
15	baseball (3)	weight (5)	buddy (10)	best (10)	soccer (5)	tennis (8)	friends (3)
16	fitness (3)	lose (5)	boxing (9)	summer (10)	make (5)	youth (8)	camps (3)
17	workout (3)	study (4)	dance (8)	play (9)	month (5)	kickers (7)	flair (3)
18	adult (3)	coach (4)	fun (8)	korean (9)	bootcamp (5)	flag (7)	events (3)
19	amateur (3)	football (4)	mens (8)	sunday (9)	baseball (5)	leagues (7)	online (3)
20	make (3)	lessons (4)	people (8)	training (9)	running (5)	lil (7)	tournament (3)

scraping craigslist

Unique Activities						
Chicago	Houston	Los Angeles	New York	Philadelphia	Phoenix	San Antonio
1 tennis (25)	workout (12)	tennis (38)	tennis (44)	tennis (25)	softball (36)	league (9)
2 female (10)	join (9)	lessons (15)	poker (30)	softball (15)	soccer (24)	tennis (9)
3 soccer (9)	trainer (8)	female (15)	female (30)	training (11)	coed (17)	basketball (7)
4 lessons (7)	training (8)	softball (14)	game (23)	coed (9)	female (16)	fitness (6)
5 softball (6)	fitness (8)	trainer (14)	lessons (19)	fitness (8)	personal (14)	coed (6)
6 play (5)	tennis (8)	training (13)	english (19)	partners (8)	training (13)	softball (4)
7 coed (5)	soccer (8)	league (13)	exchange (18)	female (8)	league (12)	training (4)
8 partners (5)	personal (8)	coed (12)	park (17)	friends (7)	football (12)	workout (4)
9 league (5)	partners (7)	workout (12)	language (17)	workout (7)	play (11)	football (4)
10 arts (4)	ladies (6)	pick (11)	softball (14)	personal (7)	summer (10)	work (4)
11 martial (4)	buddy (6)	personal (11)	central (14)	boxing (6)	indoor (10)	mens (4)
12 summer (4)	friends (6)	classes (10)	personal (12)	learn (6)	workout (10)	gym (3)
13 buddy (3)	certified (5)	soccer (10)	yoga (12)	play (6)	male (9)	like (3)
14 men (3)	activity (5)	gym (10)	club (10)	fit (6)	night (9)	advice (3)
15 baseball (3)	weight (5)	buddy (10)	best (10)	soccer (5)	tennis (8)	friends (3)
16 fitness (3)	lose (5)	boxing (9)	summer (10)	make (5)	youth (8)	camps (3)
17 workout (3)	study (4)	dance (8)	play (9)	month (5)	kickers (7)	flair (3)
18 adult (3)	coach (4)	fun (8)	korean (9)	bootcamp (5)	flag (7)	events (3)
19 amateur (3)	football (4)	mens (8)	sunday (9)	baseball (5)	leagues (7)	online (3)
20 make (3)	lessons (4)	people (8)	training (9)	running (5)	lil (7)	tournament (3)