

# PDRPy

Praca Domowa 2



## Wybrane społeczności

- **bicycles**.stackexchange.com - 56,860 postów
- **codegolf**.stackexchange.com - 176,368 postów (1.6 GB)
- **coffee**.stackexchange.com - 3,936 postów
- **crypto**.stackexchange.com - 55,368 postów
- **gamedev**.stackexchange.com - 128,190 postów
- **pets**.stackexchange.com - 18,779 postów
- **ukrainian**.stackexchange.com - 5,069 postów

**codegolf.**stackexchange.com



## **codegolf.**stackexchange.com

“Code golf is a type of recreational computer programming competition in which participants strive to achieve the shortest possible source code that implements a certain algorithm. Playing code golf is known as "golf scripting". Code golf challenges and tournaments may also be named with the programming language used (for example, Perl golf).”

- [https://en.wikipedia.org/wiki/Code\\_golf](https://en.wikipedia.org/wiki/Code_golf)

# codegolf.stackexchange.com - Zagadnienie

## MySQL, 167 characters

83

```
SELECT REPLACE(@v:='SELECT REPLACE(@v:='\ '2\ ',1+1,REPLACE(REPLACE(@v,\'\\\\\\\\\\\\\\\\\',\'\\\\\\\\\\\\\\\\\'))
```



That's right. :-)

## Brainf\*ck (755 characters)

41

This is based off of a technique developed by Erik Bosman (ejbosman at cs.vu.nl). Note that the "ESultanik's Quine!" text is actually necessary for it to be a quine!

## Prelude, 5157 4514 2348 1761 1537 664 569 535 423

69

241 214 184 178 175 169 148 142 136 133 bytes

Thanks to Sp3000 for saving 3 bytes.

These are the two shortest **Ruby** quines [from SO](#):

25

```
_"_="=p;puts _%"_";puts _%_
```

## codegolf.stackexchange.com - Założenia

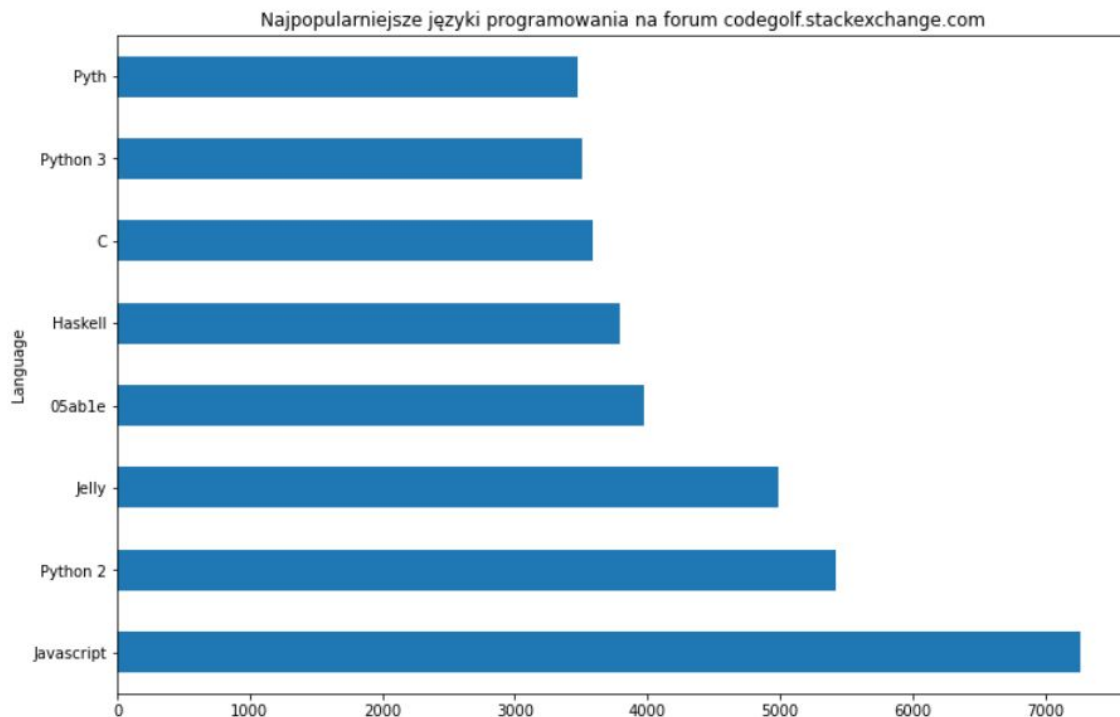
- Nie ma sztywnych zasad, opisujących co jest językiem
- Ciało odpowiedzi to HTML, więc używamy `beautifulsoup`
  - Nazwa języka zwykle znajduje się w `<h1>`, `<h2>`, lub `<h3>`
  - Usuwamy zbędne tagi (głównie przekreślenia)
- Obok nazwy zwykle znajduje się liczba opisująca długość rozwiązania
- Możemy być pewni, że liczba reprezentuje liczbę bajtów wyłącznie gdy znajduje się po niej słowo *bytes* lub podobne
- Nazwa języka znajduje się w najwyższym znalezionym wystąpieniu
- Bez NLP rozwiązania mogą opierać się jedynie na heurystykach

## codegolf.stackexchange.com - Wyniki

- 73.57% odpowiedzi zostało wziętych pod uwagę
- Wyniki częściowo zaciemnione przez niekonsekwencję w nazewnictwie
  - “Python 3” klasyfikowany jest jako inny język niż “Python 3.8”
- Typowe i popularne (*JavaScript, Python, Haskell, C...*)
- Specjalistyczne dla code golf (*Jelly, 05ab1e, Pyth, GolfScript...*)



# codegolf.stackexchange.com - Wyniki



**Tagi**

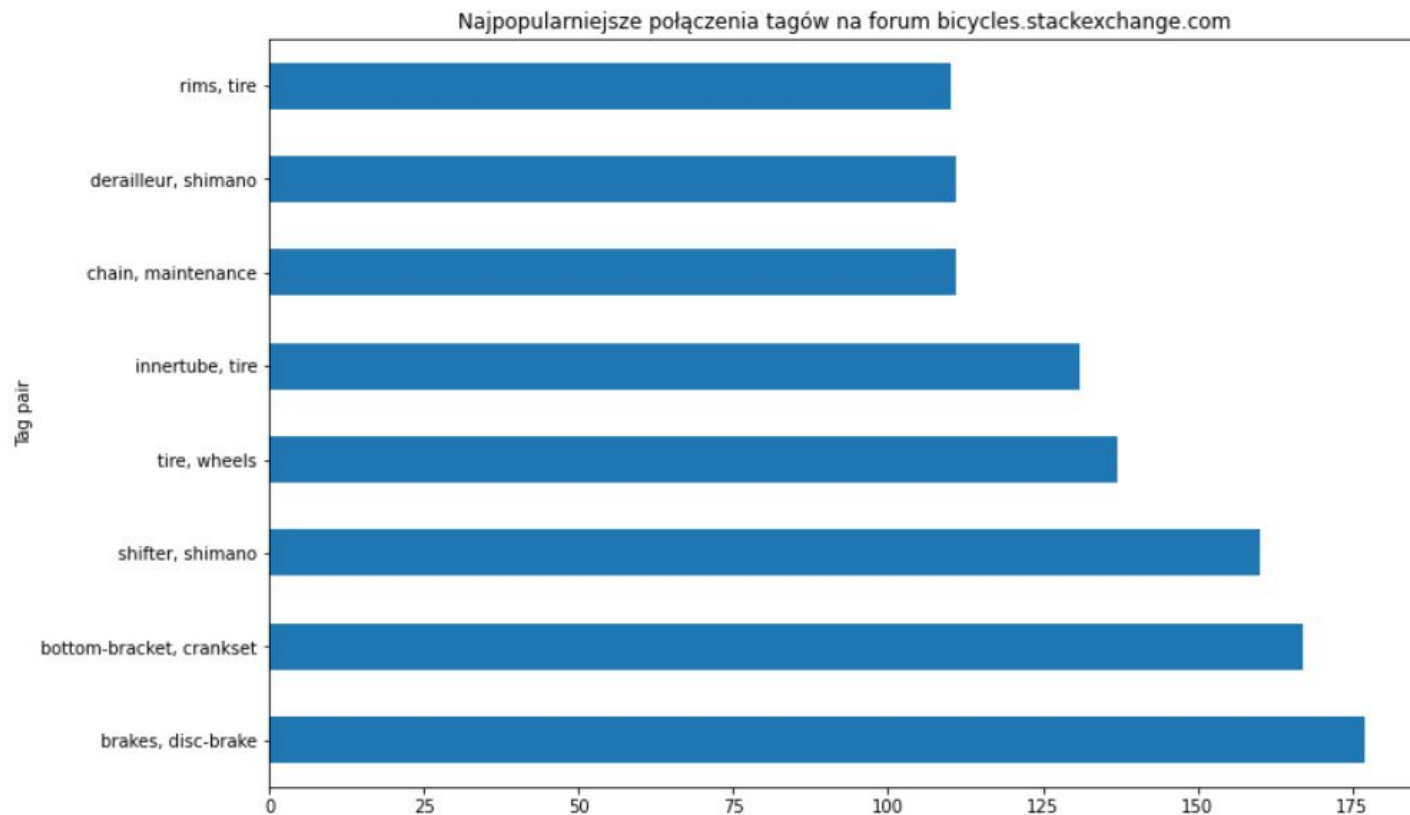


# Tagi

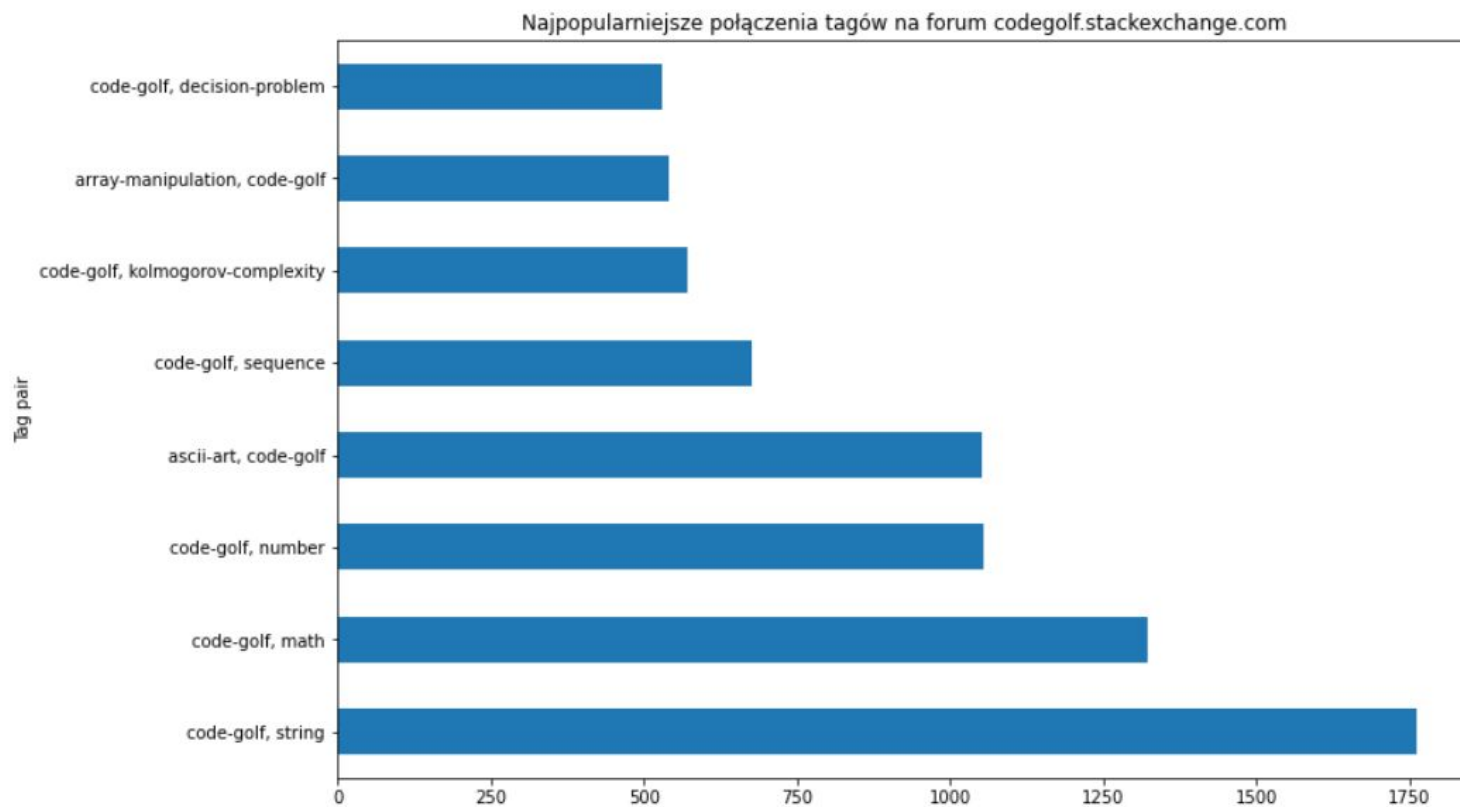
- Pierwszy napotkany problem, jaki musieliśmy rozwiązać to przetworzenie kolumny **Tags** do wydajniejszego przeszukiwania.
- Następnie wydobyliśmy wszystkie pary wspólnie występujących tagów i zliczyliśmy najczęstsze połączenia.

index		Tags
0	0	<dogs><behavior><aggression>
1	1	<dogs><training><fear><sound>
2	2	<rabbits><toilet-training>
3	3	<dogs><toilet-training>
4	6	<cats><health><safety>
...	...	...
7387	19228	<cats><behavior><training><safety><psychology>
7388	19231	<cats><house-moving>
7389	19232	<cats>
7390	19233	<fish>
7391	19236	<cats><strays><house-moving>

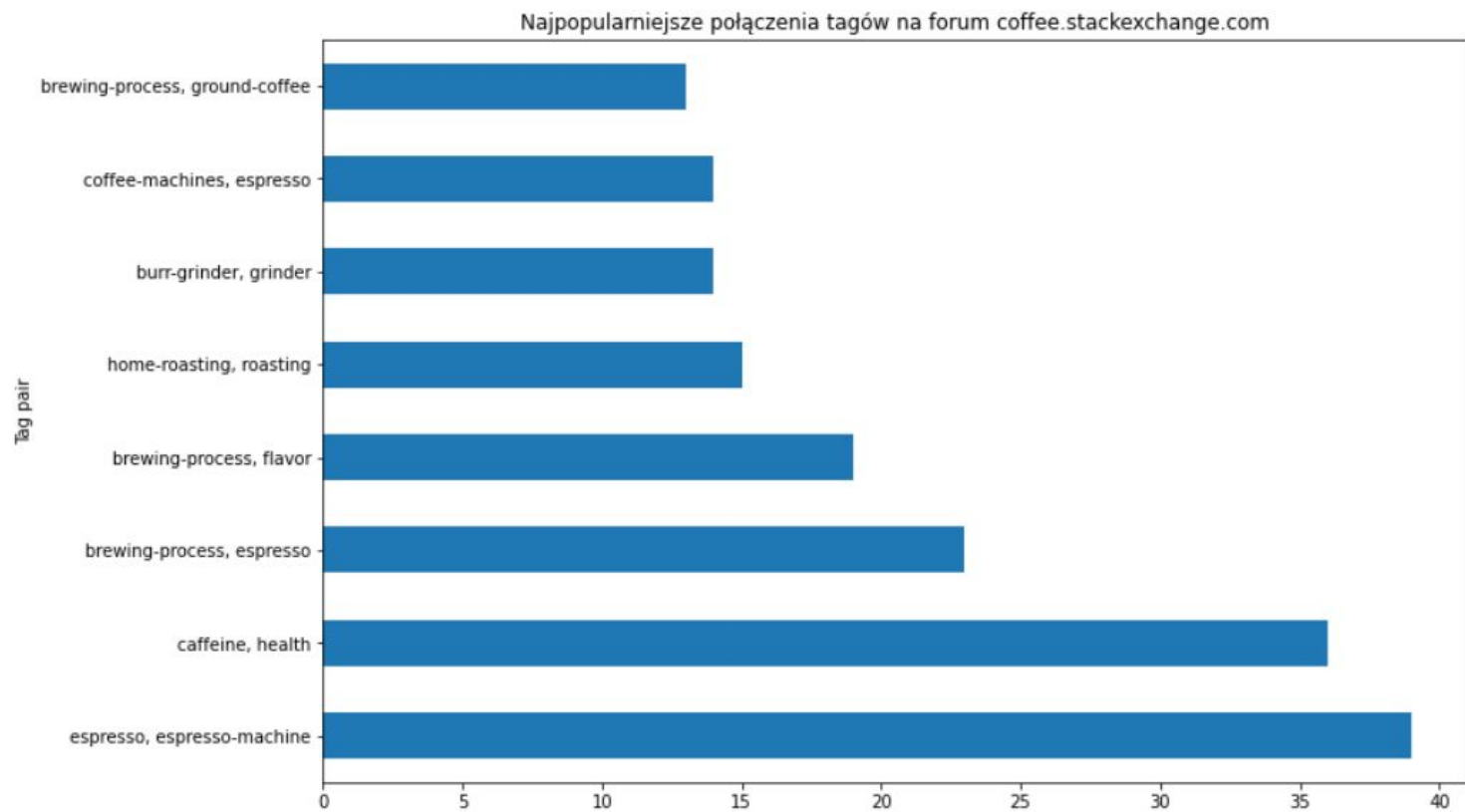
# Tagi - wyniki



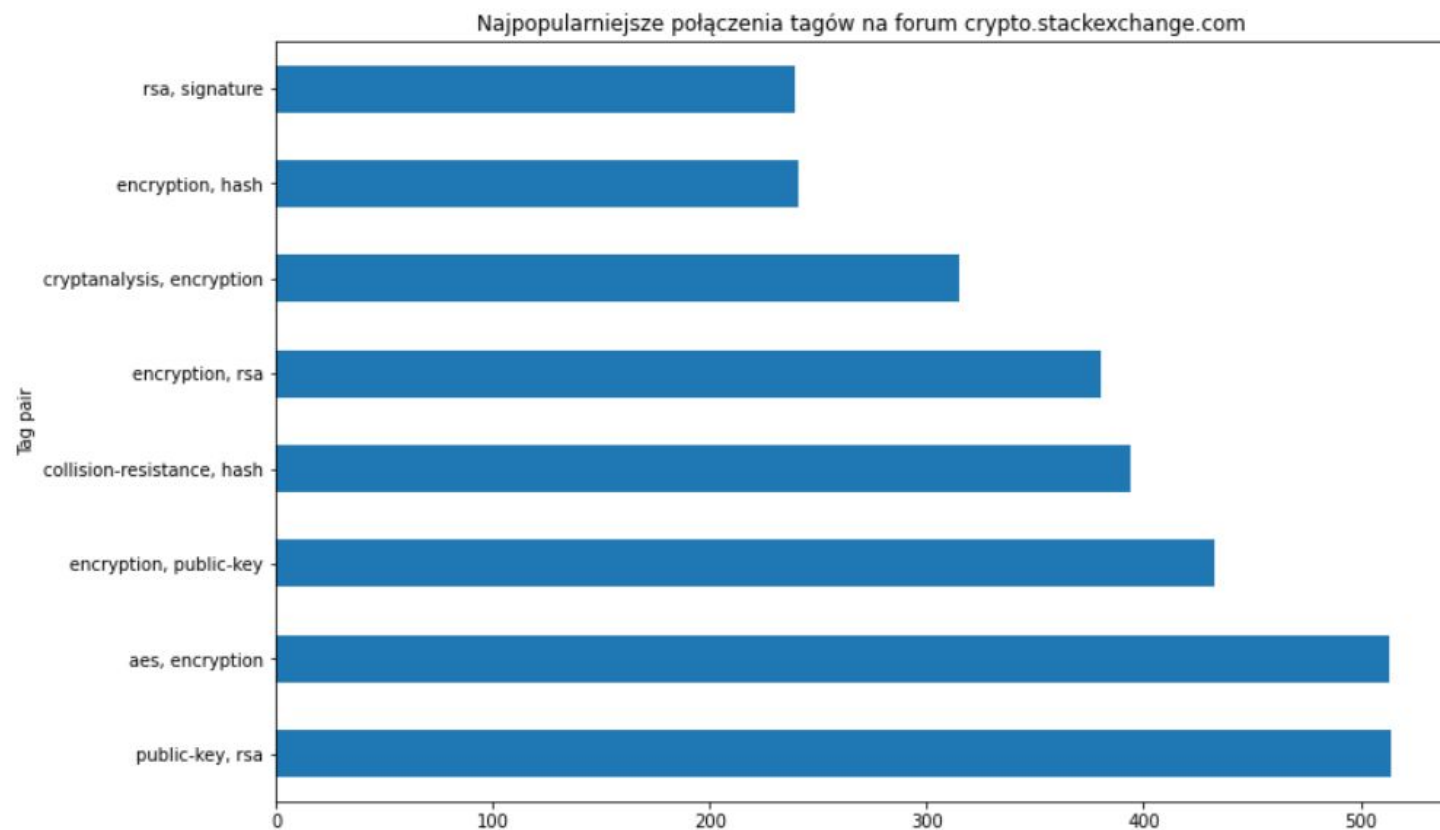
# Tagi - wyniki



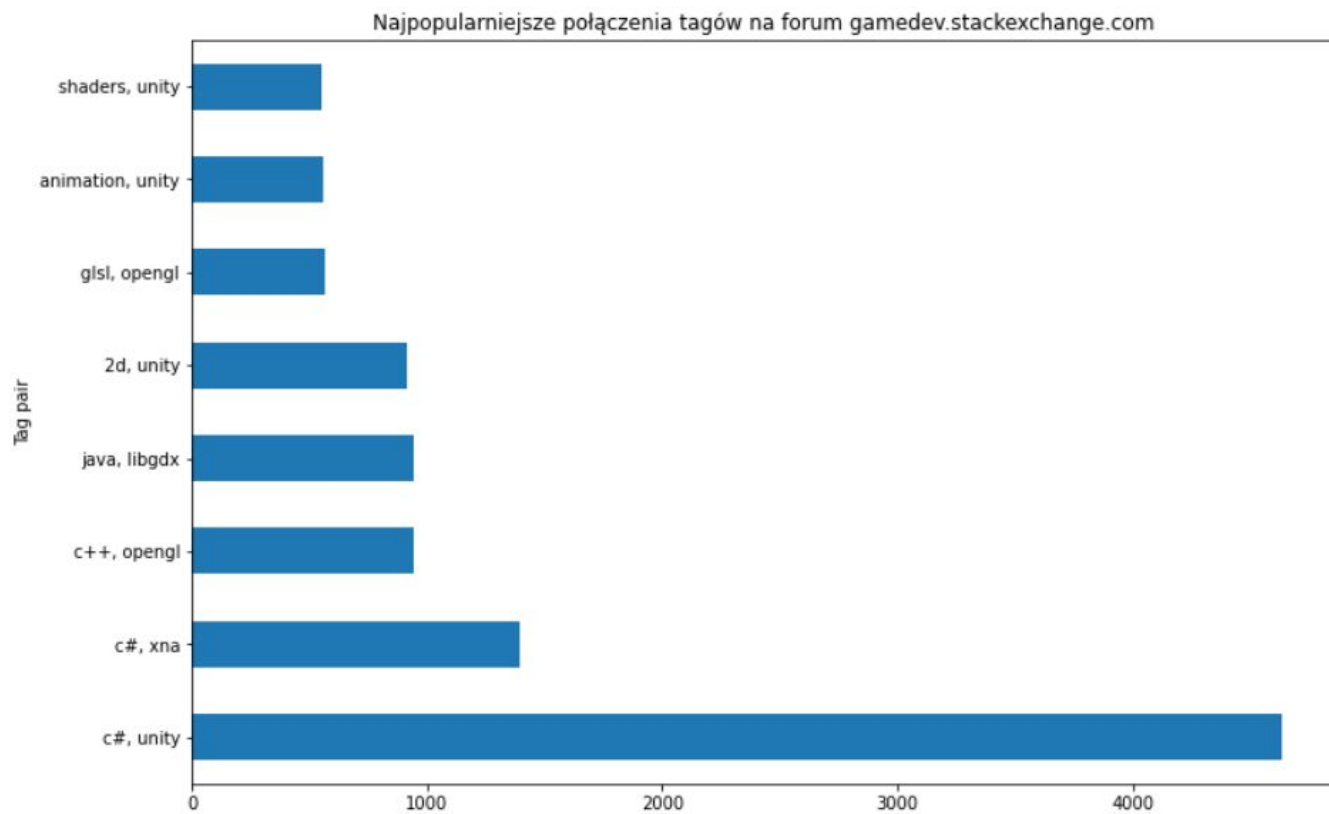
# Tagi - wyniki



# Tagi - wyniki



# Tagi - wyniki

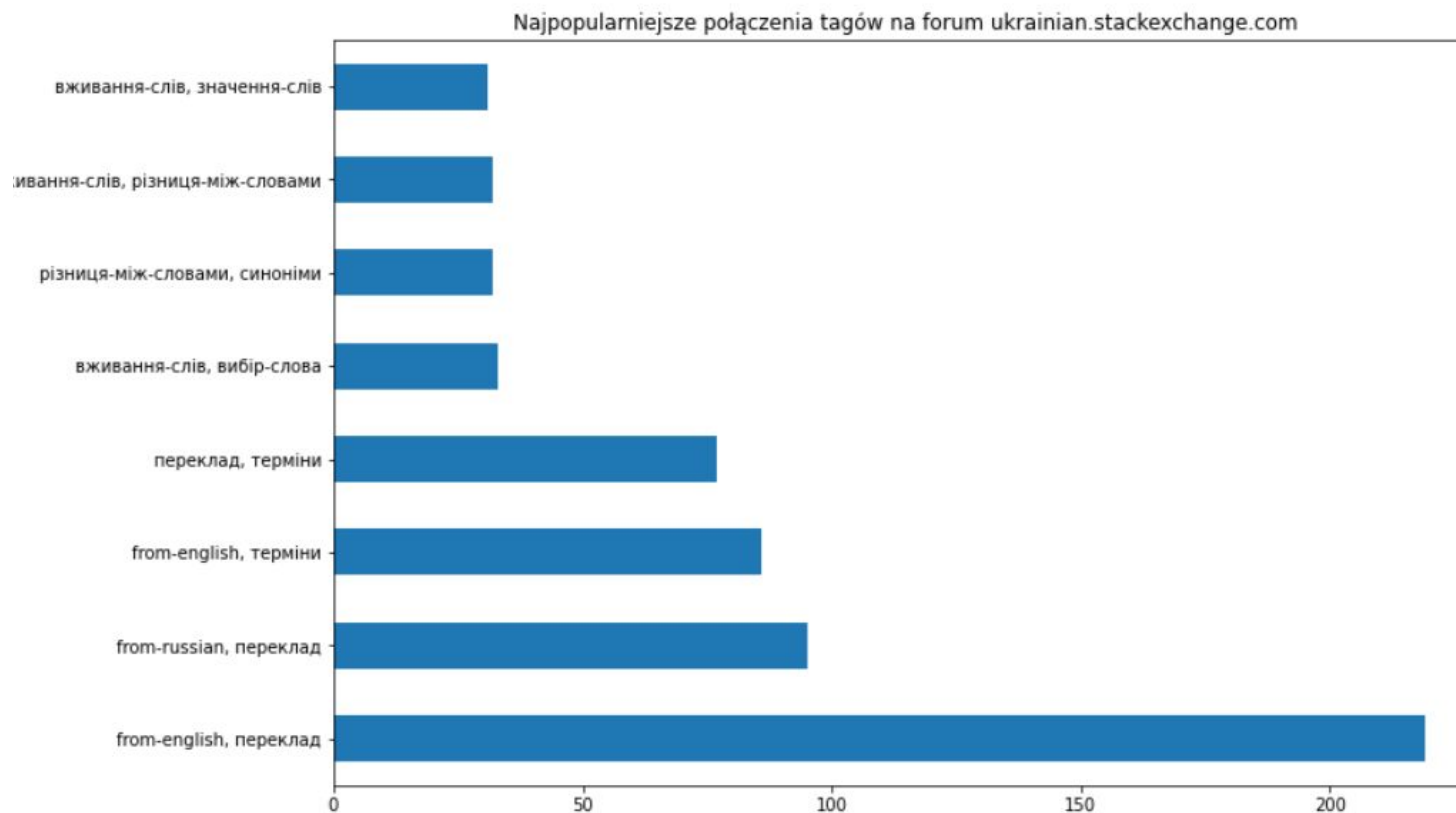




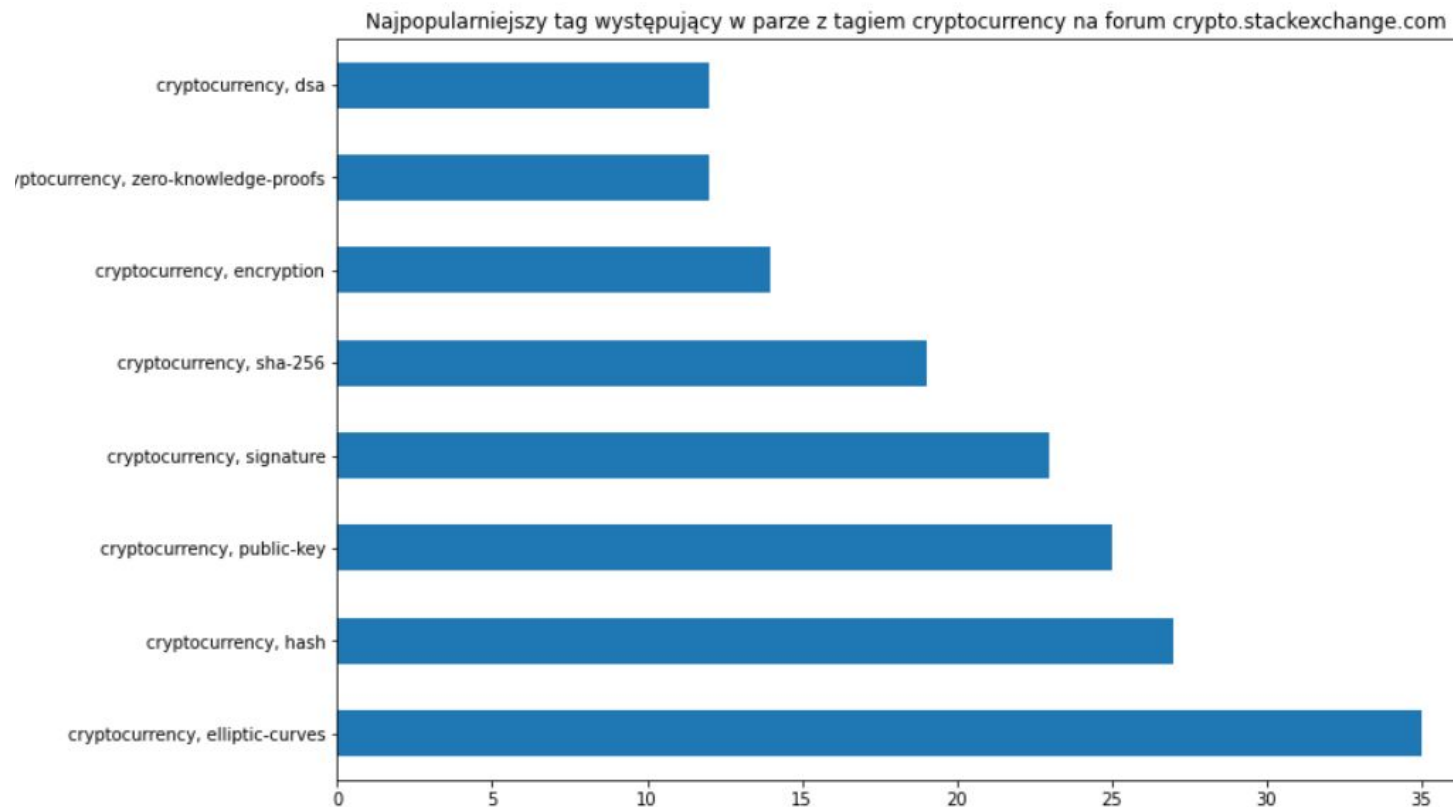
# Tagi - wyniki



# Tagi - wyniki

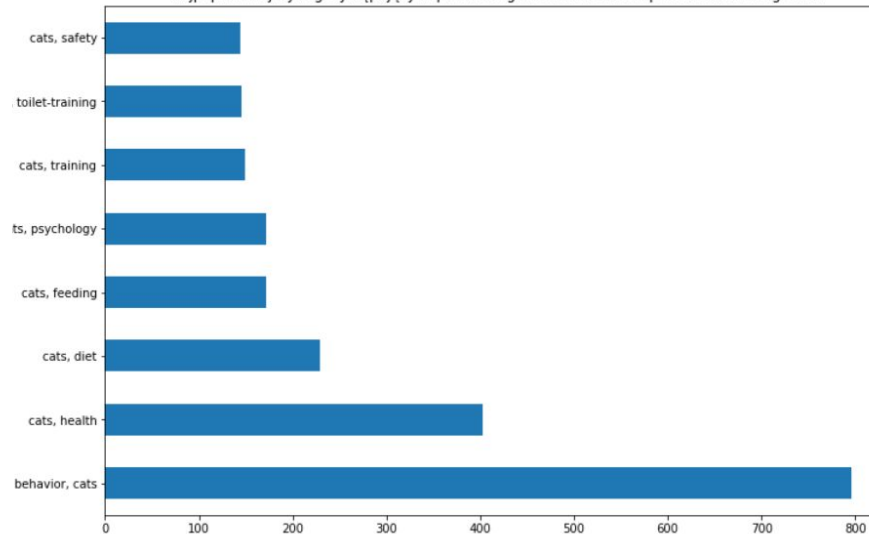


# Tagi - wyniki

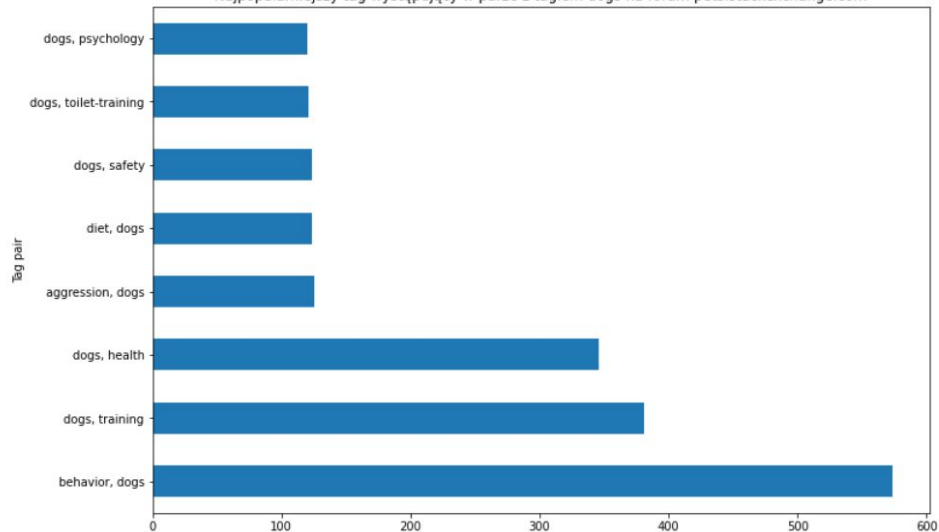


# Tagi - wyniki (*koty kontra psy*)

Najpopularniejszy tag występujący w parze z tagiem cats na forum pets.stackexchange.com



Najpopularniejszy tag występujący w parze z tagiem dogs na forum pets.stackexchange.com



**Odpowiedzi i wątki**



# Odpowiedzi i wątki

Interesują nas:

- najszybsze i najwolniejsze pierwsze odpowiedzi
- najkrócej i najdłużej trwające wątki (tzn. muszą mieć przynajmniej 1 odpowiedź)
- porównanie między społecznościami (*czy rowerzyści rzeczywiście są tacy szybcy, jak twierdzą?*)

# Odpowiedzi i wątki - wyniki (*wersja 1*)

```
In [5]: analyze_v1(bicycles_posts_df)
```

```
Out[5]:
```

	diff_time_min	diff_time_max
min	-292 days +20:16:49.880000	0 days 00:00:00
max	1619 days 12:16:06.473000	3806 days 00:28:05.467000

Otrzymaliśmy ciekawe wyniki.

Zerowe i ujemne.

```
In [6]: analyze_v1(coffee_posts_df)
```

```
Out[6]:
```

	diff_time_min	diff_time_max
min	0 days 00:00:00	0 days 00:00:00
max	1679 days 21:41:25.017000	2178 days 04:13:18.530000

Zacznijmy od ujemnych.

```
In [7]: analyze_v1(ukrainian_posts_df)
```

```
Out[7]:
```

	diff_time_min	diff_time_max
min	0 days 00:00:00	0 days 00:00:00
max	1195 days 04:04:19.527000	1429 days 06:35:39.454000

# Odpowiedzi i wątki - wyniki ujemne

	Title	CreationDate_question	CreationDate_answer	diff_time
9293	Does drafting cause resistance to the lead rider?	2012-06-26T15:56:24.383	2011-09-09T12:13:14.263	-292 days +20:16:49.880000
9293	Does drafting cause resistance to the lead rider?	2012-06-26T15:56:24.383	2011-09-09T13:09:40.740	-292 days +21:13:16.357000
9293	Does drafting cause resistance to the lead rider?	2012-06-26T15:56:24.383	2012-06-26T16:08:55.867	0 days 00:12:31.484000
9293	Does drafting cause resistance to the lead rider?	2012-06-26T15:56:24.383	2012-06-26T22:40:22.667	0 days 06:43:58.284000
9293	Does drafting cause resistance to the lead rider?	2012-06-26T15:56:24.383	2012-06-27T11:20:53.213	0 days 19:24:28.830000
9293	Does drafting cause resistance to the lead rider?	2012-06-26T15:56:24.383	2015-08-25T04:43:32.313	1154 days 12:47:07.930000

Po krótkim dochodzeniu znaleźliśmy winowajcę.



# Odpowiedzi i wątki - wyniki ujemne

Rzeczywiście użytkownik **coco**, znany dalej jako *Nostradamus* zadał (jedyne w swojej karierze na bicycles.stackexchange.com) pytanie w 2012 roku, a następnie (choćby chciałoby się powiedzieć poprzednio) użytkownicy **Daniel R Hicks** oraz **Angelo** odpowiedzieli w 2011 roku.



Dowód podróży w czasie można znaleźć dalej wiszący na forum:

<https://bicycles.stackexchange.com/questions/10069/does-drafting-cause-resistance-to-the-lead-rider>

# Odpowiedzi i wątki - wyniki zerowe

Id_question		Title	IsSame
11165	13157	I know there are better rain-forecast sources ...	True
11550	13573	What are some alternatives to fruitlessly ring...	True
12018	14102	Is the Suunto Ambit compatible with the Bontra...	True
12779	14965	What are my options for cycling mirrors?	True
13509	15827	Fixed Gear Chainring Centering	True
13606	15945	How can I run a dynamo wire from my front hub ...	True
13857	16215	How can I prevent cars from passing me too clo...	True
14217	16617	What's an "Auto-Mini" folding bike? Can I use ...	True
14377	16794	Can I carry my tandem on a bumper carrier?	True

Id_question		Title	IsSame
36	40	Який символ використовувати для позначення апо...	True
435	467	Чи є загальне правило, яке керує написанням "д...	True
473	505	Милозвучність української мови. Закони й засоби	True
600	639	Як правильно вживати слова "воєнний", "військо...	True
659	699	Чи можна вживати слово "калитка" для позначенн...	True
843	900	«Закордон» чи «за кордон», «закордоном» чи «за...	True
870	927	Чи справді вираз «ти правий» у значеннях «твоя...	True
1199	1270	Судно чи сүдно?	True
1996	3148	Лікарський чи Лікарський? Як правильно?	True

Analiza pokazała, że wszystkie odpowiedzi, napisane poniżej 1 ms po zamieszczeniu pytania należą do tego autora pytania. Ta zależność okazała się prawdziwa dla wszystkich analizowanych społeczności.

# Odpowiedzi i wątki - wyniki zerowe

## Pouring technique for drip/pour over coffee

Asked 6 years, 4 months ago   Active 5 years, 11 months ago   Viewed 657 times

asked Feb 4 '15 at 20:35



EdChum

2,177 • 13 • 31

answered Feb 4 '15 at 20:35



EdChum

2,177 • 13 • 31

---

did you immediately answer your own question? – Justin C Feb 4 '15 at 23:23

---

@JustinC yes the idea here is that this question I think will be useful for others, I was hoping for more answers which if better than my copy and paste effort I will accept – EdChum Feb 4 '15 at 23:25

---

# Odpowiedzi i wątki - wyniki (*wersja 2, tylko dodatnie*)

bicycles

	diff_time_min	diff_time_max
min	0 days 00:00:00.187000	0 days 00:01:30.724000
max	1619 days 12:16:06.473000	3806 days 00:28:05.467000

codegolf

	diff_time_min	diff_time_max
min	0 days 00:00:09.916000	0 days 00:01:14.386000
max	2286 days 01:20:07.793000	4390 days 00:56:30.434000

coffee

	diff_time_min	diff_time_max
min	0 days 00:01:04.147000	0 days 00:01:04.147000
max	1679 days 21:41:25.017000	2178 days 04:13:18.530000

crypto

	diff_time_min	diff_time_max
min	0 days 00:00:00.234000	0 days 00:00:52.530000
max	2603 days 17:38:56.997000	3426 days 18:42:12.203000

gamedev

	diff_time_min	diff_time_max
min	0 days 00:00:00.170000	0 days 00:00:09.750000
max	2976 days 22:21:36.270000	3787 days 08:41:12.380000

pets

	diff_time_min	diff_time_max
min	0 days 00:00:57.300000	0 days 00:00:57.300000
max	2227 days 12:09:37.457000	2671 days 18:58:26.343000

ukrainian

	diff_time_min	diff_time_max
min	0 days 00:00:15.744000	0 days 00:00:15.744000
max	1195 days 04:04:19.527000	1429 days 06:35:39.454000

# Dziękujemy!

Mateusz Zieliński  
Szymon Tomulewicz

