

PRZYGOTOWAŁ
PAWEŁ JAN TŁUSTY

WIESZ?

SZTUKA UZYSKIWANIA INFORMACJI



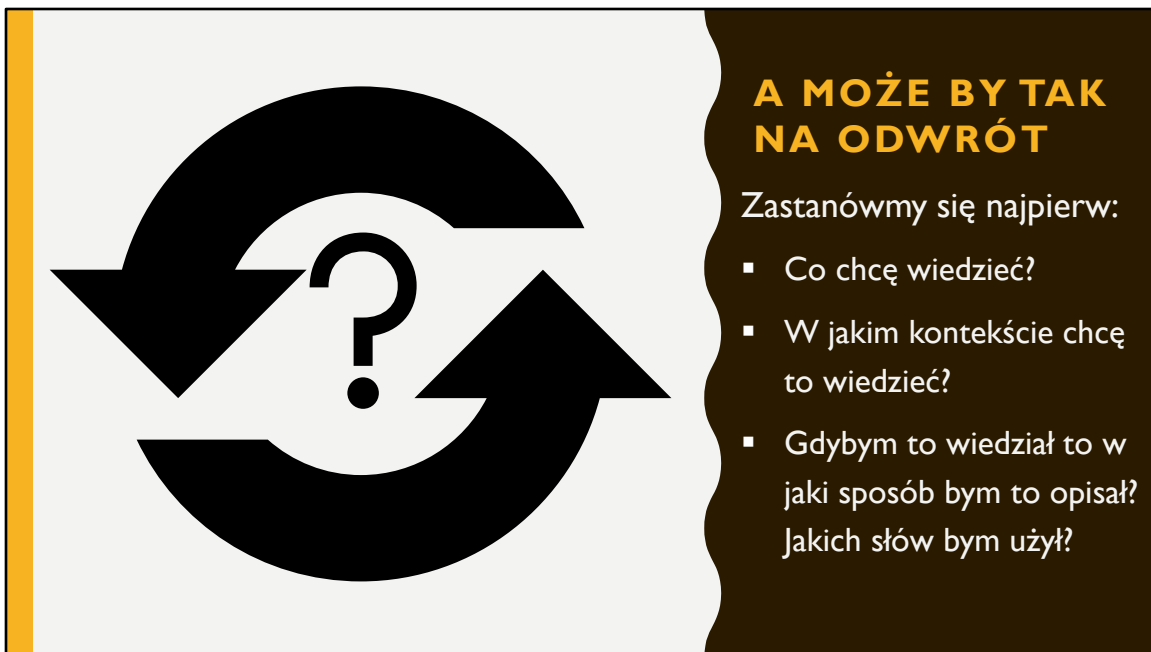
TAK TRUDNO COŚ ZNALEŹĆ?

**DLACZEGO TAK W OGÓLE CHCEMY
O TYM ROZMAWIAĆ**



CZYM TAK WŁAŚCIWIE JEST PYTANIE

Aby w ogóle zacząć zastanawiać się jak przyspieszyć swoją pracę z silnikami wyszukującymi powinniśmy się zatrzymać na chwilę i pomyśleć. Jak zadalibyśmy pytanie człowiekowi, aby dostać precyzyjną i poprawną odpowiedź?



A MOŻE BY TAK NA ODWRÓT

Zastanówmy się najpierw:

- Co chcę wiedzieć?
- W jakim kontekście chcę to wiedzieć?
- Gdybym to wiedział to w jaki sposób bym to opisał? Jakich słów bym użył?

Bo tak naprawdę to tutaj nam nie chodzi o to żeby zadać pytanie. No bo co stworzę sobie takie super pytanie, zacznę je zadawać każdemu na ulicy i nawet nie czekając na odpowiedź będę biegł do następnego przechodnia. No niby fajnie, może i mnie to w jakimś stopniu będzie cieszyć. Ale bez uzyskania jakiejś informacji raczej będzie to proces bezproduktywny

Pisząc o kontekście mam na myśli pewną przestrzeń/ płaszczyznę problemu. Nasz problem/zapytanie musimy w czymś umiejscowić (typu. Jak zrobić pętlę w C++) Mamy tu co? W jakim kontekście? Jest spora szansa na to, że się dowiemy. A gdyby tak zapytać tylko „Jak zrobić pętlę?” Otu przestrzeń nam się zmieni, ale nadal mamy jakieś szanse. To wcale nie jest błąd, to zapytanie jest poprawne, jedyne gdzie mijamy się z celem to intencje.



Nie wiem jak u Państwa ze znajomością Hiszpańskiego i Chińskiego, ale ja chyba pozostanę przy angielskim

Więcej ludzi z tym samym problemem = więcej rekordów = większa szansa na to że rozwiążemy nasz problem



W DOMU

Używając izolowanego środowiska przeczytajcie następujący artykuł.

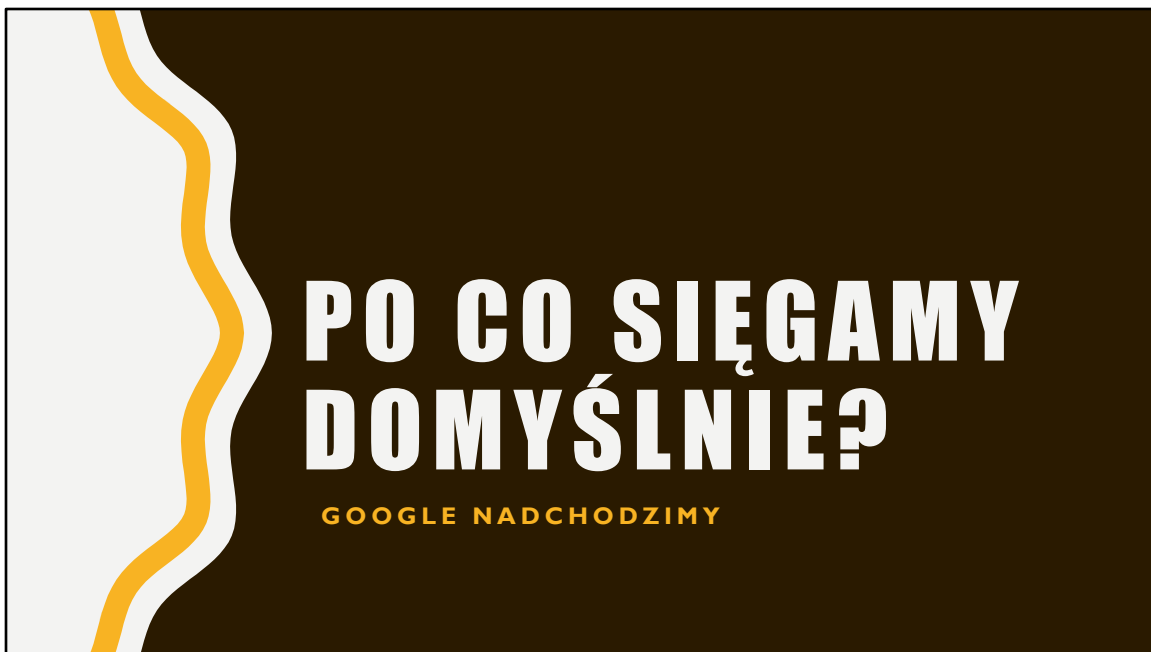
Polecam wersję angielską. Artykuł rozjaśnia wiele kwestii dotyczących uzyskiwania informacji w internecie

Link w komentarzu

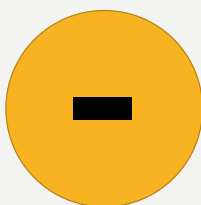
Link do artykułu:

[http://catb.org/~esr/faqs/smart-](http://catb.org/~esr/faqs/smart-questions.html?fbclid=IwAR2iRDrI3D2I2d1j8s9dsoGTQOYoz_ZKjHGvU0MQTJqW3c9LwWVRS5pNUUg)

[questions.html?fbclid=IwAR2iRDrI3D2I2d1j8s9dsoGTQOYoz_ZKjHGvU0MQTJqW3c9LwWVRS5pNUUg](http://catb.org/~esr/faqs/smart-questions.html?fbclid=IwAR2iRDrI3D2I2d1j8s9dsoGTQOYoz_ZKjHGvU0MQTJqW3c9LwWVRS5pNUUg)



KWANTYFIKATORY



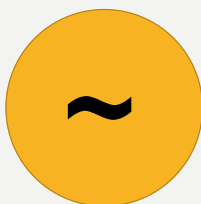
not

Wyklucza z
wyników następny
argument



AND

Logiczna
koniunkcja,
można zapisać
jako „AND”



Synonim

Poszukuje również
wyrazów
bliskoznacznych



OR

Logiczna
alternatywa,
można zapisać
jako „OR”

Zakładam, że każdy z tu obecnych wie jak wygląda wyszukiwarka Google, przez co po prostu przeskoczmy dalej.

Gdyby ktoś pytał to nie jest drukowane I tylko kreska Shiffera, ten znaczek obok backSlasha

Linkacz do zbioru większej ilości przydatnych kwantyfikatorów:

<https://ahrefs.com/blog/google-advanced-search-operators/>



Z ważniejszych rzeczy

For example, **site:youtube.com** or **site:.gov**.

Przy operatorach słownych nie ma np. site: „spacja” youtube.com, wtedy to nie zadziała spacja robi się argumentem i zwraca głupoty, albo nulla

I ta zasada się tyczy wszystkich operatorów z dwukropkiem.

Też ładnie opisane operatory wraz z przykładami.

<https://www.webnotes.com/how-to-use-shortcuts-and-symbols-in-google-search/>

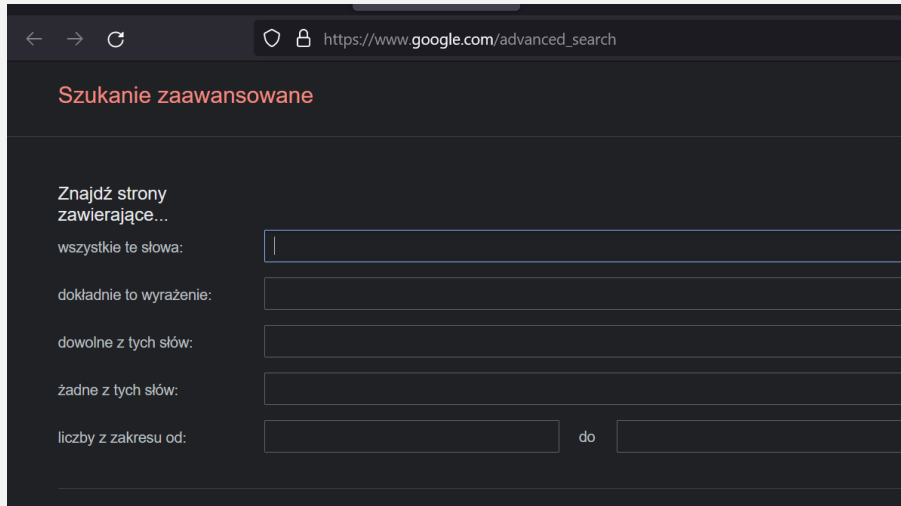
Kolejne i inne ładne opisy:

<https://support.google.com/websearch/answer/2466433?hl=en>

Filetype dokładniej:

<https://support.google.com/webmasters/answer/35287?hl=en>

ALE CZY TRZEBA TO WSZYSTKO PAMIĘTAĆ?



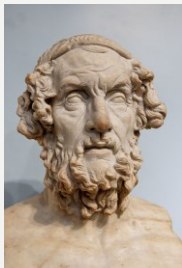
The image shows a screenshot of the Google Advanced Search page in Polish. The browser's address bar displays the URL `https://www.google.com/advanced_search`. The page title is "Szukanie zaawansowane". Below the title, there is a section titled "Znajdź strony zawierające..." (Find pages containing...). This section includes several search filters, each with a corresponding input field:

- "wszystkie te słowa:" (all these words): A single-line text input field.
- "dokładnie to wyrażenie:" (exactly this expression): A single-line text input field.
- "dowolne z tych słów:" (any of these words): A single-line text input field.
- "żadne z tych słów:" (none of these words): A single-line text input field.
- "liczby z zakresu od:" (numbers in the range from): A single-line text input field, followed by the word "do" (to), and another single-line text input field.

No niby nie, bo tu są reguły, dziwne pomysły etc. A tu mamy taki ładny graficzny interfejs. Jednakże jest on delikatnie ograniczony and trzeba się przeklikiwać. Time, time, time

Ćwiczenie

Przypadek Homera Simpsona



Używając kwantyfikatorów wyszukiwarki Google spowoduj odnalezienie niemalże wyłącznie rekordów związanych z Homerem Simpsonem nie odwołując się do jego nazwiska, typu animacji, czy czegokolwiek z nim związanego.

Psst. Użyjcie słowa „Homer”

Jest od groma ćwiczeń związanych z tworzeniem zapytań. (Nie będę podawał linków odnalezienie tego potraktujcie jako zadanie dodatkowe 😊)

Na początek można szukać ręcznie, potem można zacząć bawić się w multi-agregatory, później zagonić do tej pracy Pythona, i tak możemy rozwijać to dłuuuuuuuugo żeby nie powiedzieć, że w nieskończoność



A co w przyszłości?:

-google dorks

- Dokładniejsze przeszukiwanie portali
 - Kombajny indeksujące
 - Agregatory danych
 - Omówienie ich bezpieczeństwa
 - Dorzucenie slajdu, że używając kwantyfikatorów dostajemy więcej reklam. Trochę porad odnośnie tego.
 - Podział przeglądarek(horyzontalne, wertykalne, klastrujące). Zasady działania, bezpieczeństwo, przykłady etc.
 - Jakby się udało to proste zaprzęgnięcie pythona do robienia za nas researchu
- I zapewne coś się jeszcze wytworzy :D