#kodowanie10

Search Engines & SEO Cybersecurity

Jak działają silniki wyszukiwarek?

Co interesuje Google, Yahoo, Bing czy inne wyszukiwarki w naszej witrynie? W jaki sposób w ciągu pół sekundy miliardy stron są przeszukiwane w celu znalezienia najbardziej odpowiadających nam treści?

Dlaczego Wikipedia jest zawsze na górze listy?

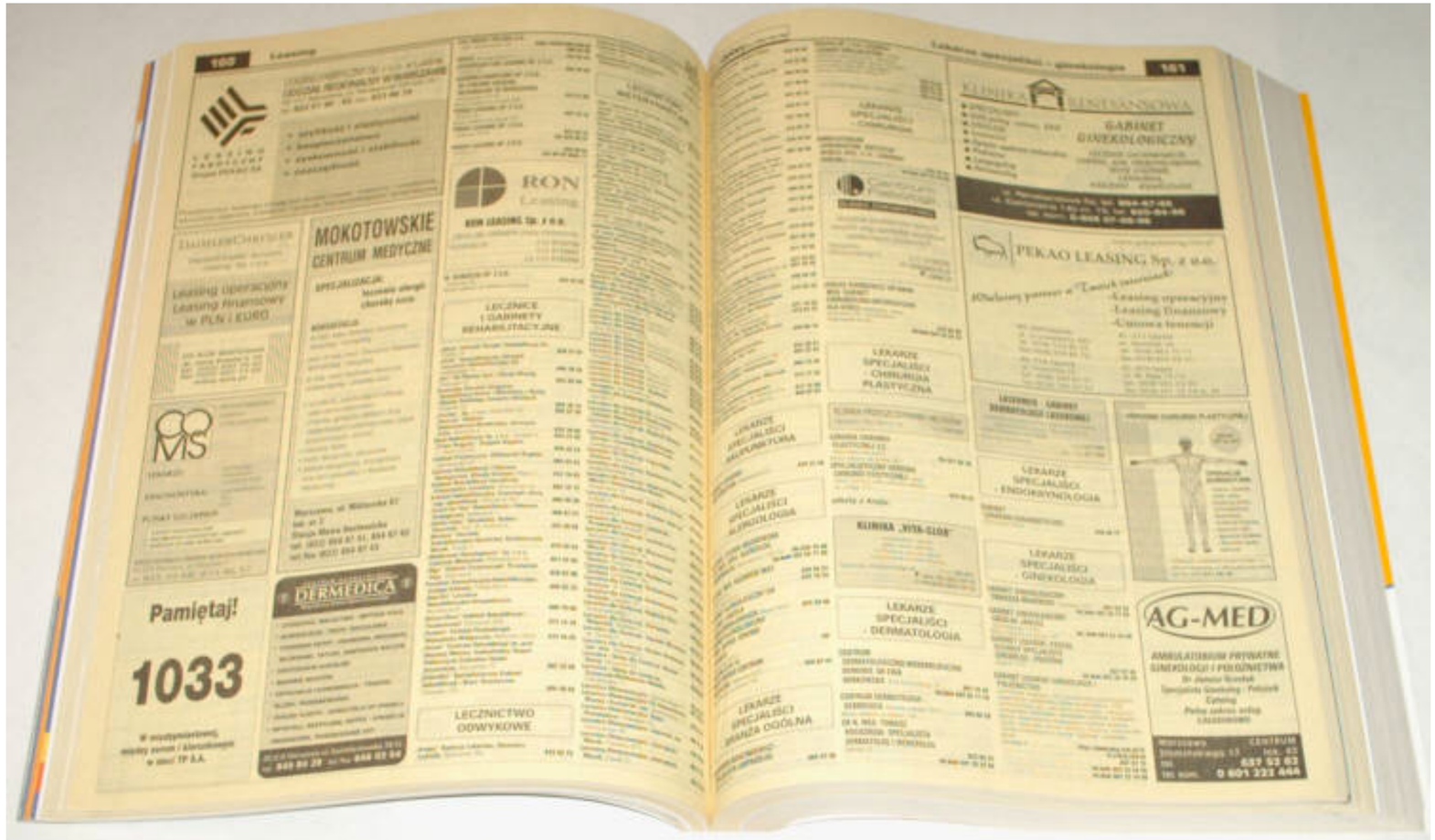
1. Wcale nie przeszukujesz sieci.

Przeszukujesz indeks wyszukiwarki.

Roboty (spiders, bots, crawlers) – małe programy – stale przeszukują sieci połączeń pomiędzy witrynami, by budować wewnętrzny katalog wyszukiwarki.

Robią to dzięki wzajemnemu odwoływaniu się stron do siebie – linków. Można też ręcznie dodać swoją stronę do katalogu, by przyspieszyć proces.

Dzięki temu podczas wyszukania nie trzeba przejrzeć witryny od deski do deski – wiadomo, co mniej więcej się w niej znajduje.



2. Pojawiają się pytania.

Wyszukiwarka znajduje setki tysięcy stron ze słowami kluczowymi zapytania.

Następnie zadaje pytania sprawdzające jak bardzo te wyniki są zbliżone do odpowiedzi na twoje potrzeby:

1. Jak często pojawia się słowo kluczowe?
2. Czy słowa są obok siebie?
3. Czy występują synonimy słów?
....
xxx. Jak istotne są miejsca występowania tych słów? (Tytuł? Treść? Komentarz?)

Jednym z superważnych pytań jest:

X. Czy ta witryna jest godna zaufania?

Tutaj pojawia się ważna rola kodera.

Źle napisana strona, strona z błędami, pozbawiona ułatwień dla niepełnosprawnych, chaotyczna lub przeładowana – to wszystko może sprawić, że boty odczytają ją jako niezaufaną. Dlatego tak ważne pisanie jest poprawnego HTML, a potem JavaScript czy PHP.

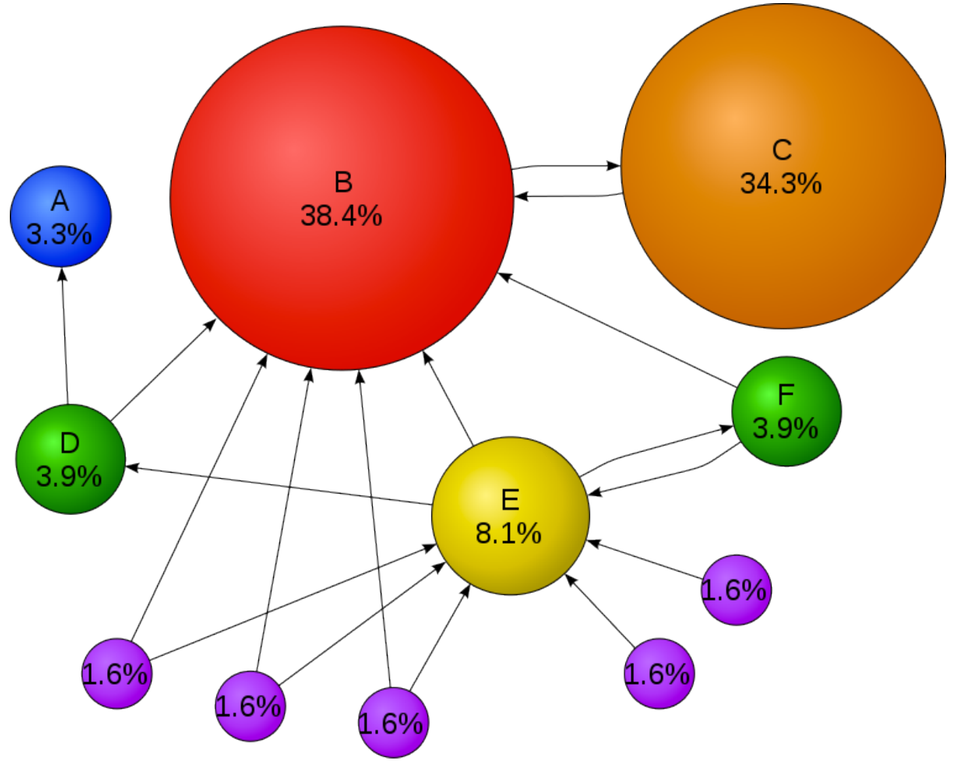
3. Pojawia się PageRank.

Twoja treść jest superważna i kluczowa – ale wielki wpływ na jej popularność ma... jej popularność.

Chociaż PageRank został oficjalnie wycofany, jego pryncypia wciąż funkcjonują.

Chodzi głównie o to, że jakość tekstu jest proporcjonalna do liczby (i jakości) tekstów, które się na niego powołują.

W praktyce oznacza to, że im więcej witryn zawiera link do naszej strony, tym ważniejsza ona jest dla wyszukiwarki.



Więcej o silnikach wyszukiwarek

Przydatne wideo:

Jak działają wyszukiwarki? (Google Team)

Jak działają wyszukiwarki? (po polsku)

Film dokumentalny o „cyfrowej rewolucji” – retro, ale częściowo bardzo aktualny

Jak poprawnie SEO.

To od ciebie zależy, czy roboty będą w stanie szybko i sprawnie znaleźć, zwiedzić i opisać twoją witrynę w katalogu.

Żadne słowa kluczowe nie załatwią sprawy, jeżeli w kodzie będzie bajzel.

1. Poprawna semantyka

Po raz tysięczny i do znudzenia: poprawny HTML eliminuje mnóstwo problemów.

– odpowiednie treści w odpowiednich tagach
– alternatywne treści dla obrazków czy wideo
– czysta struktura

Dla sprawdzenia istnieją narzędzia umożliwiające sprawdzenie, w jaki sposób naszą witrynę widzą wyszukiwarki.

2. Czysta struktura globalna

Nawigacja na naszej stronie jest ważna zarówno dla UX jak i dla robotów wyszukiwarek.

Należy z rozwagą budować drzewo strony, uważać na duplikaty treści czy puste przebiegi.

Korzystając z Wordpress możemy sobie znacznie ułatwić życie, poprawnie konstruując menu.

3. Słowa kluczowe

`<meta name=”keywords”...>` **jest absolutnie nieużyteczne i można o nim śmiało zapomnieć.**
Występowanie słów kluczowych na stronie jest ważne, a najważniejsze jest miejsce ich występowania:

– tag tytułowy
– podtytuł lub generalnie gdzieś „u góry” strony
– w atrybutach alt przy obrazach (minimum raz)
– może być zawarty w meta description tag

4. Wymuś linkowanie

Im więcej osób kliknie w link odnoszący do twojej witryny, tym większą popularność ona zdobędzie.

Fajnie sprawdzają się tutaj newslettery, social media, zachęcanie do wejścia na stronę w mejlach czy blogach – „więcej informacji znaleźć można tu: www....”.

Dodatkowe narzędzia

Nie wystarczy tylko skonstruować poprawną witrynę i wrzucić ją w otchłań sieci. Warto korzystać z dodatkowych narzędzi, które przyspieszą dziejące się i tak procesy w wyszukiwarce.

1. Mapy stron

Pojawia się tutaj kolejna wersja języka znacznikowego, zbliżonego (wizualnie) do HTML – XML.

Dobrze naszą witrynę „przepuścić” przez darmowy, dostępny online generator xml i taką mapę w xml przekazać wyszukiwarkom.

2. Robots.txt

Mówiliśmy o tagu <meta name=”robots”... Możemy również dołączyć do naszej strony cały plik tekstowy przeznaczony tylko dla robotów.

Plik taki jest oczywiście napisany w specyficzny sposób, ale szybko połapiesz się w zasadach. Treści możliwe do umieszczenia w pliku można znaleźć w sieci.

3. Webmaster Tools

Google czy Bing oferują cały zestaw narzędzi dla webmasterów, jakkolwiek głupio to nie brzmi – czasami potrzeba się nagimnastykować, żeby strona pozycjonowała się jak najlepiej, zwłaszcza na początku.

Google Webmaster tools <-link

Google magic

Płynnie przechodzimy do omówienia tego, co nam daje i zabiera firma Google.

Google (według różnych danych) obsługuje ok. 80% wyszukiwań w sieci.

80%

Google--

– przechowuje i kataloguje mnóstwo prywatnych danych
– (jak wiele serwisów i firm) alokuje w naszym komputerze mikroprogramy, które nieustannie wysyłają informacje o naszych preferencjach – tak, to cookies
– zawłaszcza mnóstwo dóbr kultury – np. Google Books
– przegląda naszą pocztę, czaty etc.
– po pewnym czasie zmusza nas do założenia sobie konta na gmail, Google Plus czy innym produkcie firmy

Google++

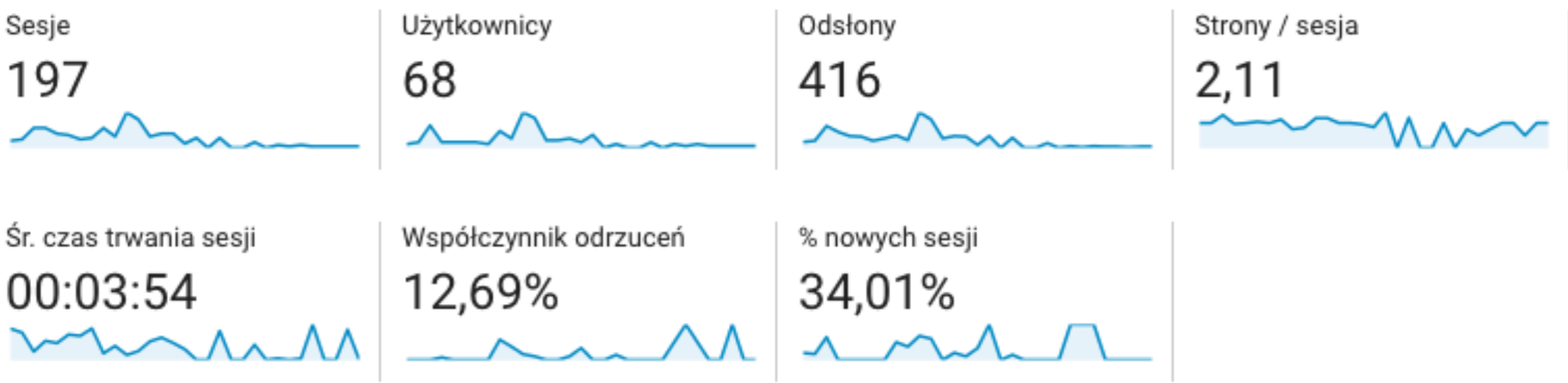
– możemy korzystać z tych skatalogowanych danych różnych osób
– cookies przyspieszają i ułatwiają obsługę witryn
– te dobra kultury (na razie) są ogólnie dostępne; dodatkowo powstają cyfrowe reprodukcje
– podglądanie naszej pracy pozwala mu dopasować treści (w tym wyniki wyszukiwania) do naszych potrzeb
– konto na Google jest wielkim ułatwieniem przy tak wielkiej popularności serwisu – klientowi pojawi się nasz adres (sklepu, studia) w mapach, godziny otwarcia, nr telefonu...

Analytics

Tak, mamy dostęp do informacji ilości osób odwiedzających naszą stronę.

Także do informacji o ich wieku, miejscu pobytu (z dokładnością do miejscowości), płci, czasie jaki poświęcili na oglądanie i w jakich godzinach robili to najchętniej.

Link do analitycs <-



A co jeśli nie chcę?

Nie wchodź do Sieci.

Niebezpieczeństwa

Możesz kontrolować ilość informacji, które o sobie umieszczasz, wyłączyć cookies, nie trzymać nagich fotek w chmurze – i tak można cię obrobić na czysto.

– jeśli masz jedno hasło do wszystkiego
– jeśli to hasło zawiera słowa występujące w słowniku
– lub logiczne ciągi numeryczne
– jeśli kompulsywnie klikasz w „kliknij mnie”
– jeśli nie czytasz między liniami
– jeśli nazbyt chętnie logujesz się do nowych serwisów
– jeśli przechowujesz wszystkie hasła w pęku kluczy a twoje hasło autoryzacji to „admin”

to lipa.

Najpopularniejsze metody

Phishing – udawanie witryny, na którą często chodzisz i się logujesz, np. portalu społecznościowego; przechwytuje twoje dane logowania, które wprowadzasz

„Metoda słownikowa” – robot wypuszcza zestaw kombinacji najpopularniejszych haseł. KTÓREŚ TRAFI.

Malware – klikasz „download” i dostajesz w prezencie robaka, którego twój antywirus nawet nie widzi – a on może zrobić absolutnie co chce, na twoim urządzeniu a nawet sieci urządzeń.

Dlaczego to w ogóle możliwe?

Oprócz naszej nieuwagi i lenistwa (hasła!) wykorzystywane są jeszcze dwa czynniki:

– brak zabezpieczeń protokołu HTTP

– brak szyfrowania danych (lub słabe szyfrowanie)

Więcej o protokołach <- link

Jak się obronić?

1. Sprawdzaj, czy w pasku adresu jest kłódeczka!

Kłódeczka wskazuje na użycie przez serwer i przeglądarkę protokołu HTTPS – bezpiecznego protokołu, który pilnuje, żeby dane nie były narażone na łatwy wyciek.

Dodatkowo twoje dane są szyfrowane, a do ich odcyfrowania potrzeba specjalnego klucza.

Więcej o szyfrowaniu <- link

2. Sprawdź czy url jest poprawny – jak coś wygląda co najmniej dziwnie, to nie klikaj.

3. Pobieraj tylko z zaufanych źródeł – to może oczywiste, ale jak potrzebujesz czegoś „na szybko”, łatwo wpaść pod tramwaj.

4. ZMIENIAJ HASŁA. UŻYWAJ SENSOWNYCH. NIE ZAPISUJ ICH NA GOOGLE DRIVE.

5. Uruchamiaj Security Updates UPEWNIAJĄC SIĘ, że powiadomienia o nich pochodzą od providera (hackerzy często wypuszczają fałszywe security update’s po to, by wrzucić ci malware)

6. Nie wysyłaj haseł PRZEZ FACEBOOK. Najlepiej nie wysyłać ich też przez e-mail, ale czasami nie ma innego wyjścia. E-mail ma większą protekcję niż messenger, który nawet nie wiadomo, kto inny czyta.

7. Używaj haseł trudnych i różnych. Są do tego specjalne programy, które kosztują $$$, ale za zamek w drzwiach też zapłacisz.

8. Kiedy się da, używaj two-factor identification. To upierdliwe potwierdzać wszystko dwa razy, ale może uratować skórę.

9. Nie uruchamiaj urządzeń od ludzi, których nie znasz. To technologiczny odpowiednik „baw się odpowiedzialnie”.

Więcej o cybersecurity <- link