

Najwięcej dobrych stron

Liczba dokumentów w polskim Internecie rośnie w astronomicznym tempie. Szacuje się, że rocznie przybywać będzie ponad 120 tysięcy nowych polskich domen¹. Dodatkowo, istniejące już serwisy znacznie zwiększają swoją objętość. Chcąc umożliwić polskim Użytkownikom łatwe przeszukiwanie rosnących zasobów Internetu NetSprint trzykrotnie zwiększył ilość indeksowanych stron. Ich liczba osiągnęła rekordowy poziom 67.000.000 i nadal rośnie.

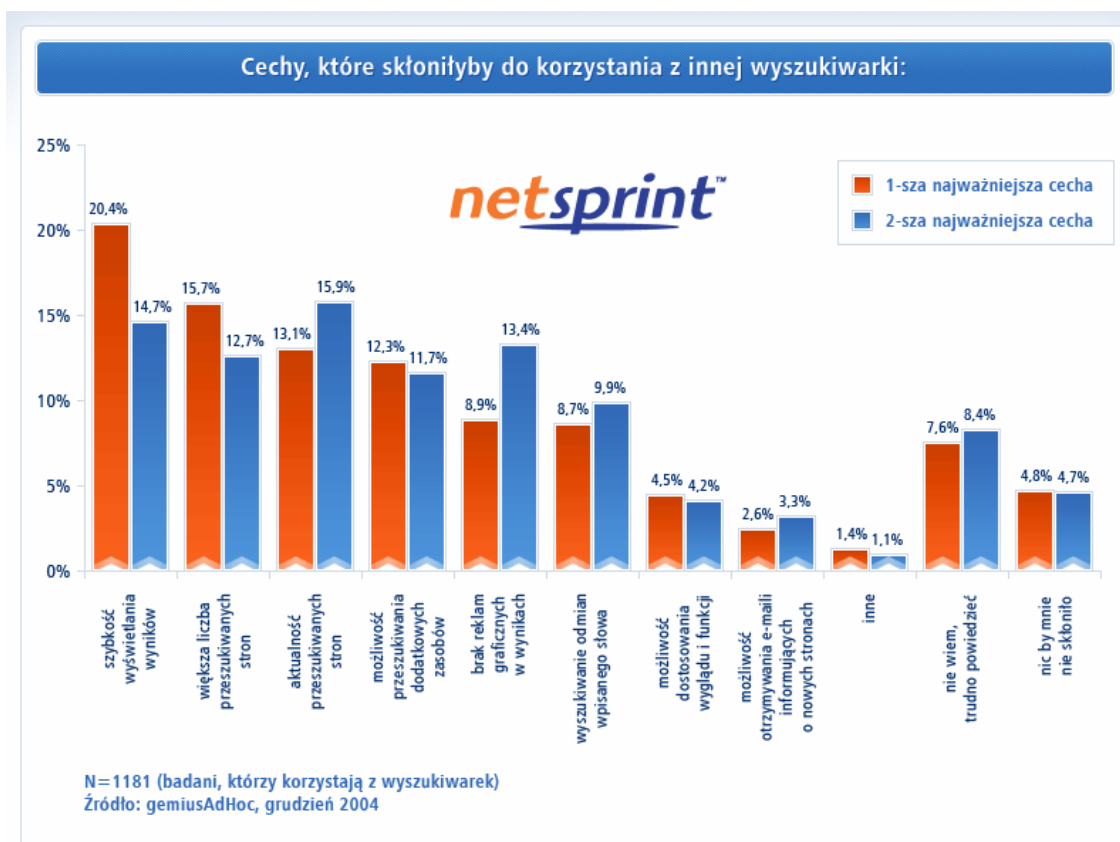
Internet rozwija się bardzo dynamicznie. W wielu obszarach wiedzy dostępne w nim zasoby przekraczają już teraz te zgromadzone w największych bibliotekach. Według sierpniowego raportu zaprezentowanego w serwisie statystykidomen.pl aktualna liczba polskich domen internetowych wynosi 343 643, przy czym w ciągu ostatnich 12 miesięcy przybyło ich ponad 92 000. Największe polskie portale osiągają rozmiary kilkuset tysięcy dokumentów.

Dogonić Internet

W tym kontekście jedną z najważniejszych funkcji wyszukiwarki staje się efektywne przeszukiwanie rosnących zasobów. Wg badania przeprowadzonego przez Gemius S.A. na zlecenie NetSprint.pl Internauci przyznali, że jest to dla nich - zaraz po szybkości - najważniejsza funkcjonalność wyszukiwarki. Aż 28% wskazało liczbę przeszukiwanych stron jako cechę, która skłoniłaby ich do zmiany wyszukiwarki na inną.

„Obserwując dynamiczny wzrost liczby stron w sieci oraz bazując na przeprowadzonym badaniu w styczniu 2005 r. podjęliśmy wspólnie z Wirtualną Polską decyzję, aby przeszukiwać więcej dokumentów” – mówi Artur Banach, prezes Zarządu NetSprinta. Wszystkie polskie wyszukiwarki przeszukiwały w tamtym okresie około 20 mln dokumentów. Na podstawie naszych analiz określiliśmy, że, aby zaspokoić potrzeby Użytkowników i „dogonić” polski Internet powinniśmy przeszukiwać we wrześniu więcej niż 50 mln polskich stron.

¹ Więcej informacji na: <http://statystykidomen.pl/NASKvsRejestratorzy/>



Najwięcej dobrych stron

„Użytkownicy chcą przeszukiwać nie tylko jak największą liczbę stron. Równie ważne jest dla nich, aby były to dokumenty zawierające wartościowe treści” – mówi Piotr Kozłowski – szef zespołu IT w NetSprincie.

„Naszym celem było, aby nasza wyszukiwarka:

- przeszukiwała we wrześniu powyżej 50 mln polskich dokumentów i później wzrastała w tempie szybszym od polskiego Internetu
- pobierała i indeksowała codziennie najbardziej opiniotwórcze serwisy polskiej sieci (dzięki czemu wartościowe dokumenty będą mogły być przeszukiwane przez polskich Użytkowników już 24 h po ich publikacji).
- lepiej od innych wyszukiwarek niwelowała znaczenie spamu w wynikach wyszukiwania (gwarantuje to wyższą jakość prezentowanych wyników).
- prezentowała precyzyjne wyniki już na pierwszej stronie wynikowej.”

- wylicza Piotr Kozłowski.

W wyniku podjętych w okresie styczeń-wrzesień działań:

- czterokrotnie wzrosła platforma sprzętowo-technologiczna;
- wydajność robota zbierającego (tzw. spidera) wzrosła 5-krotnie umożliwiając pobieranie 6 mln dokumentów z sieci dziennie;
- stworzono unikalny algorytm Traffic Index, dzięki któremu pomimo zwiększenia liczby dokumentów precyzyjne wyniki znajdują się na pierwszej stronie wynikowej²;
- zaprojektowano zmiany w serwisie netsprint.pl, który dzięki nowym funkcjom i wyglądowi podkreśla korzyści płynące z wprowadzonych rozwiązań technologicznych;
- ilość indeksowanych stron wzrosła do 67 mln i nadal rośnie.

Walczymy ze spamem

Wzrost ilości indeksowanych stron wiąże się z koniecznością wzmożenia walki ze spamem. Bez działań tego rodzaju zwiększenie liczby zbieranych dokumentów nie przyniosłoby istotnych korzyści dla Użytkowników.

NetSprint wspólnie z Wirtualną Polską zawsze przywiązywał bardzo dużą wagę do tego, by prezentowane wyniki nie wprowadzały Internautów w błąd. Walka z takimi działaniami polegała na udoskonaleniu algorytmu oceny stron i wzmożonej pracy zespołu usuwającego korzystające z nielegalnych technik witryny.

Już w 2003 roku NetSprint wspólnie z portalem WP opublikował zasady swojej polityki zwalczania spamu:

http://firma.netsprint.pl/index.php?option=com_content&task=view&id=19&Itemid=173

Podkreśleniem wagi jaką wszystkie polskie wyszukiwarki przywiązują do tego zagadnienia był kolejny, przyjęty przez grupę roboczą SEO przy IAB Polska Kodeks Etyczny Pozycjonowania. Warto podkreślić, że w tej inicjatywie uczestniczyli również przedstawiciele portalu Onet.pl

http://firma.netsprint.pl/index.php?option=com_content&task=blogcategory&id=94&Itemid=183

W 2005 roku NetSprint systematycznie nasilał działania w zakresie walki ze spamem. Wprowadzono szereg narzędzi automatycznie ograniczających pozycje nieuczciwie pozycjonowanych stron. W procesie tym zasadnicze znaczenie miało wprowadzenie innowacyjnego algorytmu Traffic Index. Ponadto nieuczciwe działania spamerów są surowo karane, z eliminacją serwisu z indeksu wyszukiwarki włącznie.

² więcej informacji na ten temat w dokumencie Traffic Index

Przykładem sukcesów NetSprinta w tym obszarze jest tekst opublikowany w Magazynie InternetStandard „Porno w Kutnie.pl”. Dotyczy on kłopotów jakie z polskimi spamerami ma wyszukiwarka Google.com. Dzięki swojej polityce NetSprint.pl uniknął podobnych problemów .

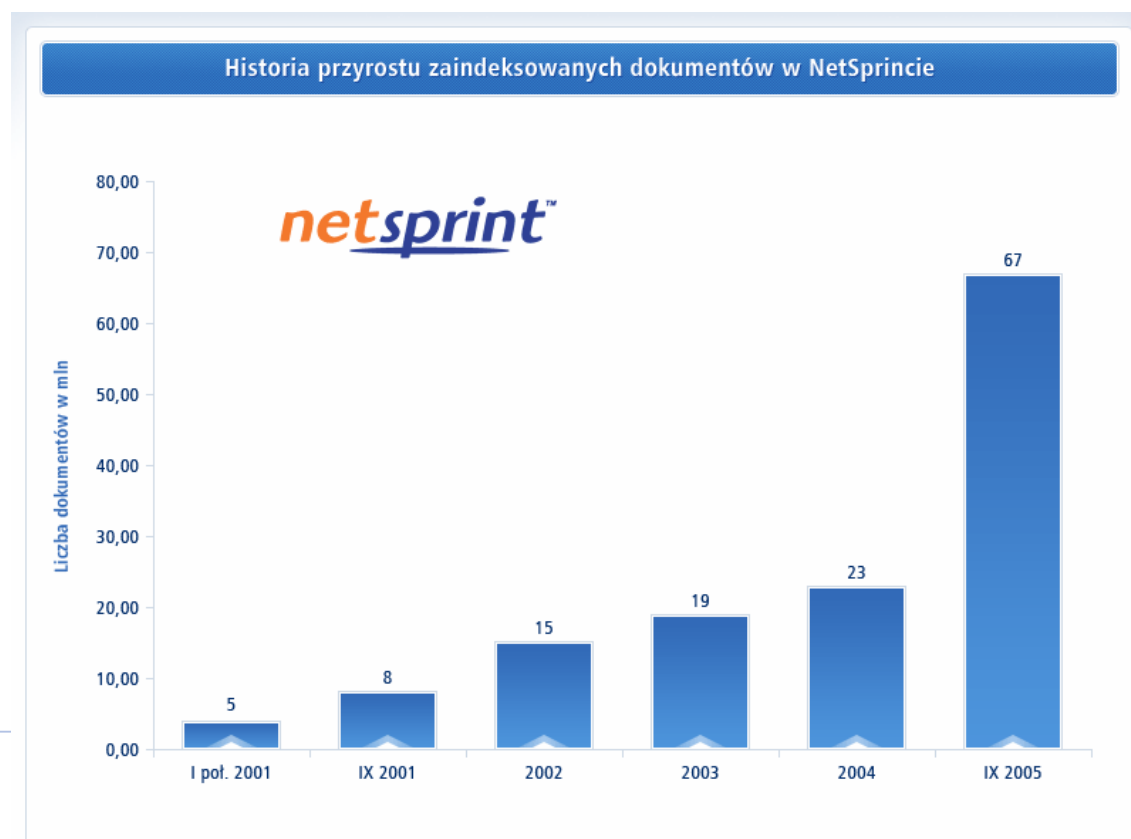
<http://www.internetstandard.pl/news/82690.html>

NetSprint rośnie

Od początku działalności strategicznym celem firmy był stały wzrost liczby przeszukiwanych stron. Pierwsze wdrożenie wyszukiwarki NetSprint miało miejsce w 2000 roku na portalu Hoga.pl. NetSprint już wówczas indeksował 4 mln dokumentów. O standardzie tego rozwiązania świadczy fakt, że w rankingu PC World Komputer w lutym 2001 roku NetSprint został uznany za najlepszą wyszukiwarkę w Polsce.

We wrześniu 2001 roku wystartował serwis www.netsprint.pl. W chwili startu umożliwił przeszukiwanie **8 milionów** dokumentów (niewielu internautów pamięta, że z tego rozwiązania korzystał m.in. portal Arena.pl.) Natomiast w połowie 2002 roku NetSprint niemalże podwoił swój indeks osiągając **15 mln stron www**.

W grudniu 2002 roku miało miejsce kolejne przełomowe wydarzenie - premiera NetSprinta na portalu Wirtualna Polska. Współpraca ta wiązała się nie tylko z intensywnym przyspieszeniem prac rozwojowych, ale również konsekwentnym wzrostem liczby stron przeszukiwanych zarówno na portalu wp.pl jak i netsprint.pl (są to dwie osobne instancje). W roku 2003 NetSprint indeksował średnio 19 mln stron, zaś rok później **23 miliony**.



Komputery i biblioteki

Na koniec chcielibyśmy nieco z przymrużeniem oka dokonać porównania wielkości zasobów przeszukiwanych przez NetSprinta z największymi polskimi bibliotekami. Już obecnie liczba dokumentów dostępnych w NetSprintcie wielokrotnie przewyższa liczbę woluminów w Bibliotece Narodowej czy Bibliotece Jagiellońskiej. Oczywiście, zestawienie nie uwzględnia faktu, że średnia wielkość dokumentu bibliotecznego jest z całą pewnością większa niż przeciętna strona www. Biorąc jednak pod uwagę wcześniejsze informacje dotyczące tempa wzrostu polskiego Internetu oczywiste jest, że wkrótce Internet będzie stanowił doskonałą alternatywę dla największych polskich bibliotek. I to dostępną w czasie poniżej 1 sekundy.....

