

Patent Google o SEO i pozycjonowaniu...



W naszych artykułach o SEO i pozycjonowaniu zawsze staramy się czerpać z najlepszych źródeł... Z jednej strony podążamy za poszlakami poszukując odpowiedzi na domniemany [wpływ ruchu na wyniki organiczne](#), staramy się rozszyfrować [wpływ CTR na nasze pozycje w Google](#)... Z drugiej strony zasięgamy informacji u źródła przeprowadzając [wywiady z Guglarzem](#) lub tłumacząc zachodnie [webinary z Googlersami](#)...

W komentarzach pod artykułami czasami narzekacie że [Guglarz nic ciekawego powiedzieć nie może](#) lub że [niektóre artykuły to bardziej fantastyka](#) niż rzeczywistość... Przestaliśmy zatem pytać a zaczęliśmy szukać... Mamy nadzieję że dzisiejszy artykuł zadowoli naszych nawet najbardziej wybrednych czytelników.

Dotarliśmy do bardzo wiarygodnego źródła informacji jakim są naszym zdaniem **patenty Google!** Kładą one zupełnie nowe światło na naszą dotychczasową wiedzę... Żadna książka nie tłumaczy tak dobrze działania największej na świecie wyszukiwarki jak jej własne, opisane i opatentowane rozwiązania...

- Jak Google określa **datę powstania dokumentu** ?
- Jak **aktualizacje treści** wpływają na pozycję strony ?
- Czy i jak Google mierzy **szybkość przyrostu linków** ?

Tych informacji nie znajdziecie w plikach pomocy Google... żaden **forumowy Guru SEO** nie odpowie wam na te pytania...

Odpowiedzi przedstawiamy w dzisiejszym artykule...

"Information retrieval based on historical data"

Taką nazwę nosi pierwszy patent, który chcemy dzisiaj omówić... Pod patentem podpisał się m.in. Matt Cutts, który co ważne jest z Google od początków istnienia tej firmy.

Myślę że warto osobiście zapoznać się z całą treścią [patentu Google](#)... Artykuł jest bardzo obszerny, w sumie około 20 stron tekstu A4 napisanego czcionką 12. Jeszcze dzisiaj wgramy tu PDF'a do druku.

Data Powstania dokumentu

Ważne z punktu widzenia zrozumienia zasad określania np. Duplicate Content.

Dział: DESCRIPTION - EXEMPLARY HISTORY DATA

Akapy: [0034] - [0044]

"...there are several techniques that can be used to determine a document's inception date..."

Jest kilka technik mierzenia daty powstania dokumentu...

Kilka najważniejszych wymienionych w dokumencie:

- Data pierwszego zaindeksowania
- Data odkrycia pierwszego linku prowadzącego do dokumentu (jak istnieje link to i strona musi już istnieć)
- W niektórych przypadkach za datę powstania dokumentu może zostać uznana **data rejestracji domeny** na której powstał dokument.
- Data powstania dokumentu zapisana przez serwer czyli np. nagłówek "*HTTP If-Modified-Since*" itp.

Do określenia daty powstania dokumentu Google może używać kombinacji kilku technik.

"Search engine may use the inception date of a document for scoring of the document. For example, it may be assumed that a document with a fairly recent inception date will not have a significant number of links from other documents (i.e., back links). For existing link-based scoring techniques that score based on the number of links to/from a document, this recent document may be scored lower than an older document that has a larger number of links (e.g., back links). When the inception date of the documents are considered, however, the scores of the documents may be modified (either positively or negatively) based on the documents' inception dates."

Do oceny dokumentu wyszukiwarka może wykorzystać datę jego powstania, na przykład, można śmiało założyć, iż dokument z dość niedawną datą powstania nie będzie miał znacznej liczby linków z innych dokumentów. Dla aktualnych algorytmów opartych na ocenie linków do i z **dokumentów**, młodszy dokument może więc zostać oceniony niżej od starszego, który

liczbę takich linków ma znacznie większą. Gdy jednak data powstania dokumentów jest brana pod uwagę przy ustalaniu rankingu, wówczas wyniki wyszukiwania mogą być modyfikowane (pozytywnie lub negatywnie) na podstawie tychże dat.

“Consider the example of a document with an inception date of yesterday that is referenced by 10 back links. This document may be scored higher by search engine than a document with an inception date of 10 years ago that is referenced by 100 back links because the rate of link growth for the former is relatively higher than the latter. While a spiky rate of growth in the number of back links may be a factor used by search engine 125 to score documents, it may also signal an attempt to spam search engine 125. Accordingly, in this situation, search engine 125 may actually lower the score of a document(s) to reduce the effect of spamming.”

Rozważmy przykład dokumentu z wczorajszą datą powstania, który posiada 10 linków. Ten dokument może być oceniany wyżej niż dokument z datą powstania 10 lat temu, który jest wskazywany przez 100 linków ponieważ szybkość wzrostu linków nowego dokumentu jest wyższa. Podczas gdy szybki przyrost linków może być czynnikiem ocenionym na plus dokumentu, może być to jednak także sygnał o próbie SPAMu. W związku z tym zbyt szybki przyrost linków może skutkować obniżeniem oceny dokumentu.

“...search engine may use the inception date of a document to determine a rate at which links to the document are created (e.g., as an average per unit time based on the number of links created since the inception date or some window in that period). This rate can then be used to score the document, for example, giving more weight to documents to which links are generated more often.”

...wyszukiwarka może użyć daty powstania dokumentu do oceny szybkości przyrostu liczby linków (średnia ilość linków utworzona w określonym czasie od powstania dokumentu). Ten współczynnik może następnie być użyty do oceny dokumentu np. mocniej punktując dokumenty do których linki są tworzone częściej [lub bardziej regularnie].

Zależność można podsumować poniższym wzorem:

$$H=L/\log(F+2),$$

H - historia przyrostu linków

L - ilość linków prowadzących do dokumentu

F - czas, który minął od utworzenia dokumentu

“For some queries, older documents may be more favorable than newer ones. [...]search engine 125 may determine the age of each of the documents in a result set (e.g., using their inception dates), determine the average age of the documents, and modify the scores of the documents (either positively or negatively) based on a difference between the documents' age and the average age.”

Dla niektórych zapytań, starsze dokumenty mogą być bardziej faworyzowane niż nowe. [...] wyszukiwarka może określić wiek każdego z dokumentów, wyliczyć średni wiek wszystkich dokumentów z wyników wyszukiwań, a także modyfikować Score Rank dokumentów (znów pozytywnie lub negatywnie) w oparciu o różnicę między wiekiem poszczególnych dokumentów a średnim wiekiem wszystkich stron.

Kontent - Uaktualnienia i Zmiany

Dział: DESCRIPTION - Content Updates/Changes

Akapity: [0046] - [0056]

Ten punkt dobrze wyjaśnia dlaczego regularnie aktualizowane strony i blogi tak ładnie wchodzą w SERPy.

"...information relating to a manner in which a document's content changes over time may be used to generate (or alter) a score associated with that document. For example, a document whose content is edited often may be scored differently than a document whose content remains static over time. Also, a document having a relatively large amount of its content updated over time might be scored differently than a document having a relatively small amount of its content updated over time."

...informacje odnoszące się do sposobu, w jaki treści dokumentu ulega zmianie z biegiem czasu mogą zostać wykorzystane do generowania (lub zmiany) Score Rank owego dokumentu. Na przykład, dokument, którego treść jest często edytowana może być oceniany inaczej niż dokument, którego treść pozostaje niezmienna. Ponadto, dokument o stosunkowo dużej ilości aktualizowanej treści w miarę upływu czasu może być punktowany inaczej niż dokument o stosunkowo małej ilości aktualizowanej treści.

Wzór:

$$U=f(UF, UA)$$

f - suma lub suma ważona

UF - współczynnik częstotliwości zmian

UA - procentowa ilość zmienianej treści w określonym czasie

"UA may also be determined as a function of one or more factors, such as the number of "new" or unique pages associated with a document over a period of time. Another factor might include the ratio of the number of new or unique pages associated with a document over a period of time versus the total number of pages associated with that document. Yet another factor may include the amount that the document is updated over one or more periods of time (e.g., n % of a document's visible content may change over a period t (e.g., last m months)), which might be an average value. A further factor might include the amount that the document (or page) has changed in one or more periods of time (e.g., within the last x days)."

“According to one exemplary implementation, UA may be determined as a function of differently weighted portions of document content. For instance, content deemed to be unimportant if updated/changed, such as Javascript, comments, advertisements, navigational elements, boilerplate material, or date/time tags, may be given relatively little weight or even ignored altogether when determining UA. On the other hand, content deemed to be important if updated/changed (e.g., more often, more recently, more extensively, etc.), such as the title or anchor text associated with the forward links, could be given more weight than changes to other content when determining UA.”

UA może być również określana jako funkcja jednego lub więcej czynników, takich jak:

- liczba “nowych” lub unikalnych stron związanych z dokumentem powstałych przez pewien okres czasu
- stosunek liczby nowych lub unikalnych stron związanych z dokumentem w porównaniu do całkowitej liczby stron pozwiązanych [tematycznie] z tym dokumentem
- wartość średnią stopnia aktualizacji dokumentu w ciągu jednego lub większej ilości okresów czasu (np. n % z widocznej treści dokumentu może ulec zmianie w okresie t (np. ostatnie m miesięcy))
- liczba zmian treści danego dokumentu w jednym lub większej ilości okresów czasu (np. w ciągu ostatnich x dni).

Według jednego z zastosowań, UA może być określona jako funkcja inaczej ważonych części treści dokumentu. Na przykład, zmiana zawartości może przy ustalaniu UA zostać uznana za mało ważną (lub nawet zignorowana) jeżeli dotyczy takich kwestii jak:

1. kod JavaScript,
2. komentarze,
3. reklamy,
4. elementy nawigacyjne,
5. formularze,
6. znaczniki daty i czasu

Z drugiej strony, zawartość uznaje się za ważną, jeżeli aktualizacje i zmiany (np. częstsze, świeższe, obszerniejsze, itp.) dotyczą takich elementów jak tytuł tekstu lub anchor text linków wychodzących, te elementy mogą mieć większą wagę przy określaniu UA.

...wszystko to może mieć stosunkowo małą wagę lub być nawet całkowicie ignorowane przy określaniu UA. Z drugiej strony, zawartość uznaje się za ważną, jeżeli aktualizacje i zmiany (np. częstsze, świeższe, obszerniejsze, itp.) dotyczą takich elementów jak **tytuł tekstu** lub **anchor text** linków wychodzących, te elementy mogą mieć większą wagę przy określaniu UA.

“UF and UA may be used in other ways to influence the score assigned to a document. For example, the rate of change in a current time period can be compared to the rate of change in another (e.g., previous) time period to determine whether there is an

acceleration or deceleration trend. Documents for which there is an increase in the rate of change might be scored higher than those documents for which there is a steady rate of change, even if that rate of change is relatively high. The amount of change may also be a factor in this scoring."

UF i UA mogą być wykorzystywane również do wywierania innego rodzaju wpływu na wynik przypisany do dokumentu. Na przykład, wskaźnik zmian w bieżącym okresie czasu można porównać do tempa zmian w innym (np. poprzednim), aby ustalić, czy istnieje przyspieszenie czy spowolnienie trendu. Dokumenty, dla których istnieje wzrost wskaźnika zmian mogą być wyżej niż te, które mają stałe tempo zmian, nawet jeżeli wskaźnik zmian jest stosunkowo wysoki. Wielkości zmian mogą również być czynnikiem w tej punktacji.

"In some situations, data storage resources may be insufficient to store the documents when monitoring the documents for content changes. In this case, search engine may store representations of the documents and monitor these representations for changes. For example, search engine may store "signatures" of documents instead of the (entire) documents themselves to detect changes to document content. In this case, search engine may store a term vector for a document (or page) and monitor it for relatively large changes. According to another implementation, search engine may store and monitor a relatively small portion (e.g., a few terms) of the documents that are determined to be important or the most frequently occurring (excluding "stop words")."

W niektórych sytuacjach, zasoby przechowywania danych mogą być niewystarczające do przechowywania dokumentów w celu kontroli zmian treści. W takim przypadku wyszukiwarka może przechowywać reprezentacje dokumentów i oświadczenia do monitorowania tych zmian, na przykład, "cechy charakterystyczne" [sumy kontrolne ?] danych dokumentów. W takich warunkach wyszukiwarka może przechowywać wektor słów kluczowych dokumentu (lub strony) i monitorować go w momentach stosunkowo dużych zmian [supplemental index ?]. Zgodnie z innym zastosowaniem, wyszukiwarka może przechowywać i monitorować stosunkowo niewielką część (np. kilka słów kluczowych) z dokumentów, które są uznane za ważne lub najczęściej występujące (z wyłączeniem "stop words").

Ostatnie fragmenty powyższego akapitu nawiązują do jednego z pytań zadanych Guglarzowi podczas naszego wywiadu:

Cezary Lech: - Czy Google indeksuje tylko część treści na stronach znajdujących się w suplementalu? Podobno Google indeksuje tylko tzw. "ważne" słowa stron znajdujących się w Supplemental Index?

„Kaspar Szymanski: W przeszłości Supplemental index był używany do nietypowych dokumentów, które pojawiały się na raczej niecodziennie zapytania. Z perspektywy webmastera, niższa frekwencja pobierania była główną różnicą pomiędzy Supplemental indeksem a naszym głównym indeksem. Inżynierowie z Google pracowali intensywnie nad rozwiązaniem tego problemu i zniwelowaniem różnic między indeksami, odnosząc sukces we wrześniu 2007. W chwili obecnej różnice są

minimalne. Każdemu, kto by chciał wiedzieć więcej o tym, jak Google indeksuje i rankuje strony, polecam ten link :-)"

Wracając do dalszych zapisów patentu...

"For some queries, documents with content that has not recently changed may be more favorable than documents with content that has recently changed. [...] In other words, search engine may determine a date when the content of each of the documents in a result set last changed, determine the average date of change for the documents, and modify the scores of the documents (either positively or negatively) based on a difference between the documents' date-of-change and the average date-of-change."

Dla niektórych zapytań, dokumenty, których treść ostatnio się nie zmieniła mogą mieć bardziej korzystne pozycje niż dokumenty, których treść uległa zmianie. [...] Innymi słowy, wyszukiwarka może określić ostatnią datę zmian w zawartości każdego z dokumentów w wynikach wyszukiwania, ustalić średnią datę zmian w dokumentach i zmodyfikować ocenę dokumentu (pozytywnie lub negatywnie) bazując na różnicy pomiędzy datą ostatniej zmiany a średnią datą ostatniej zmiany dla wszystkich dokumentów.

Reasumując:

Dla niektórych zapytań treść stron z niezmienną się zawartością może być bardziej wartościowa niż treść często aktualizowana.

Analiza zapytań

Dział: DESCRIPTION - Query Analysis

Akapity: [0058] - [0065]

Zgodnie z założeniami patentu, istnieje nawet kilka algorytmów odpowiadających za dostosowanie wyników wyszukiwania do zapytania użytkownika np.:

"...one query-based factor may relate to the extent to which a document is selected over time when the document is included in a set of search results. In this case, search engine might score documents selected relatively more often/increasingly by users higher than other documents."

- Google może podwyższać pozycje dokumentów które są częściej wybierane przez użytkowników.

"...terms relating to a "hot" topic that is gaining/has gained popularity or a breaking news event would conceivably appear frequently over a period of time. In this case, search engine may score documents associated with these search terms (or queries) higher than documents not associated with these terms."

"...further query-based factor may relate to a change over time in the number of search results generated by similar queries. A significant increase in the number of search results generated by similar queries, for example, might indicate a hot topic or breaking news and cause search engine to increase the scores of documents related to such queries. "

- ten fragment może być wskazówką do niedawnych [testów Jeża](#)... który starał się wprowadzić nazwę firmy w miejsce sugerowanych i powiązanych z zapytaniem fraz.

"...factor may relate to queries that remain relatively constant over time but lead to results that change over time. For example, a query relating to "world series champion" leads to search results that change over time (e.g., documents relating to a particular team dominate search results in a given year or time of year). This change can be monitored and used to score documents accordingly. "

Streszczając: Wyniki jakie należy prezentować użytkownikowi na zapytania typu "world series champion" muszą się zmieniać w czasie wraz ze zmianą stanu faktycznego czyli np. zwycięscy rozgrywek sportowych w danym roku.

Kilka fragmentów w których Google przyznają że ruch i CTR może mieć wpływ na wyniki wyszukiwań...

*"...another query-based factor might relate to the "staleness" of documents returned as search results. The staleness of a document may be based on factors, such as document creation date, anchor growth, **traffic**, content change, forward/back link growth, etc. "*

Wyraźnie wymieniony, szeroko pojęty **ruch** obok tak oczywistych czynników jak: wiek, anchor linków czy aktualizacja treści...

"For some queries, recent documents are very important (e.g., if searching for Frequently Asked Questions (FAQ) files, the most recent version would be highly desirable). [...] More specifically, search engine may consider how often users favor a more recent document that is ranked lower than an older document in the search results. "

Dla części zapytań aktualność dokumentów może być szczególnie ważna (np. szukając działu FAQ "Najczęściej zadawanych pytań" aktualność będzie bardzo istotna). Wyszukiwarka może zwracać uwagę na to jak często bardziej aktualne dokumenty znajdujące się niżej w wynikach są częściej wybierane niż starsze dokumenty.

"In some situations, a stale document may be considered more favorable than more recent documents. [...] For example, if for a given query, users over time tend to select a lower ranked, relatively stale, document over a higher ranked, relatively recent

document, this may be used by search engine as an indication to adjust a score of the stale document. “

W niektórych sytuacjach starszy dokument może być bardziej faworyzowany niż ten aktualny. Na przykład, jeśli dla danego zapytania użytkownicy zdają się częściej wybierać stronę starszą i niżej notowaną niż aktualną stronę z czołówki, to może to zostać użyte przez wyszukiwarke i może mieć wpływ na pozycje (Score Rank) dokumentu.

Tym samym Google potwierdza w tym patencie możliwy wpływ wyborów dokonywanych przez użytkowników na pozycje stron w wynikach wyszukiwań.

W dalszych częściach patentu Google znajdujemy jeszcze bardziej precyzyjne informacje... wybaczenie ale przestanę cytować gdyż artykuł rozrósł się już do monstrualnych rozmiarów... Dalej czyste sparafrazowane tłumaczenie:

Czynniki związane z linkowaniem

Dział: DESCRIPTION - Link-Based Criteria

Akapity: [0067] - [0080]

Google jest w stanie mierzyć wiele czynników wpływających na ocenę dokumentu i związanych z linkami, są to m.in.:

- Daty pojawiania się nowych linków do witryny/podstrony
- Daty, w których linki do witryny/podstrony zniknęły
- Zmieniające się w czasie zachowanie linków do danej strony oraz wszelkie powstałe w ten sposób trendy i "tendencje" np. czy strona w ogólnym rozrachunku ma więcej nowo powstałych linków, czy wręcz przeciwnie. Tendencja wskazująca na spadek ilości nowych linków może wskazywać, iż dany dokument jest „nieaktualny”, natomiast tendencja wzrostowa może wskazywać na jego potencjalną „świeżość”.
- Google może sprawdzić ilość nowych linków do dokumentu powstałych w określonym czasie i porównać ją do ilości linków, które dokument otrzymał od daty swojego powstania. Może również ustalić wiek najstarszego z y% ostatnio stworzonych linków i porównać go do wieku pierwszego stworzonego linku”.
- Google w patencie podaje przykład dwóch stron stworzonych 100 dni wcześniej:
 - Strona 1 – 10% linków powstało w ciągu ostatnich 10 dni
 - Strona 2 – 0% linków powstało w ciągu ostatnich 10 dni
 - Te dane mogą również zostać użyte żeby „przewidzieć czy dana strona przynależy do konkretnego typu (np. stron, które nie są już aktualizowane, stron z rosnącą lub malejącą popularnością, lub do stron zastąpionych itp.)”

Ocena świeżości samego linku może również zostać użyta przy ustalaniu pozycji stron/podstron. Kilka czynników może wpłynąć na „świeżość” linku:

- Data pojawienia się
- Data zmiany anchor tekstu
- Data zmiany na stronie, na której znajduje się link
- Data pojawienia się strony, na której znajduje się link

- Google informuje, iż według nich link, który pozostaje niezmienny kiedy strona jest (w znaczącym stopniu) aktualizowana jest linkiem „dobrym i istotnym”

Inne czynniki oceniające link to:

- Zaufanie do linku (Google w szczególności wspomina dokumenty rządowe jako te z przypisanym zwiększonym zaufaniem (np. domeny .gov))
- Jaki autorytet i wiarygodność (Trust Rank) posiadają strony linkujące
- Świeżość strony/podstrony – wspominają stronę domową Yahoo! jako tą gdzie linki często znikają i pojawiają się.
- „Suma mocy (wagi) linków” kierujących do strony/podstrony może zostać użyta do zmiany pozycji. Google zmierzy świeżość strony na podstawie świeżości prowadzących do niej linków oraz świeżości stron, z których te linki pochodzą.
- Rozkład linków w czasie również zostanie zmierzony, przeglądarka oceni łączną ilość powstałych linków do strony/podstrony w określonym czasie oraz to, kiedy one powstawały.

Google może użyć daty pojawienia się linku do „wykrywania spamu”, „gdy właściciele dokumentów lub ich koledzy tworzą linki do swoich własnych stron w celu uzyskania lepszych pozycji w wyszukiwarce”. Google twierdzi, iż na prawidłowych stronach/podstronach „linki zwrotne powstają stosunkowo wolno” i że „nagły wzrost liczby backlinków” może oznaczać albo „gorący temat” albo „próbę spamu”.

Google podaje stronę CDC po wybuchu paniki wokół SARS jako przykład „gorącego tematu”

Google wspomina 3 przykłady spammerskiego pozyskiwania linków – „wymianę linków”, „kupowanie linków” oraz „pozyskiwanie linków od dokumentów które nie posiadają jasno określonej ‘polityki linkowania’ „(na przykład księgi gości, referrery i „strony pozwalające dobrowolnie dodawać linki do dokumentu”)

Spadek liczby linków, które posiada dokument w określonym czasie może zostać użyty aby wskazać jego nieistotność i Google zaznacza, że w takim przypadku wyszukiwarka będzie przypisywać mniejszą wartość linkom wychodzącym z takich stron lub nawet je ignorować.

Dynamika (rotacyjność) zmian linków również jest mierzona i oceniana na podstawie tego jak konsekwentnie i systematycznie przypisywane są linki do danej strony. Google podaje przykład „wyróżnionych linków dnia” i zaznacza, linki takie także będą liczone i dokument będzie otrzymywał Score Rank wyznaczony na podstawie wszystkich linków zwrotnych (także tych rotujących). - "Google Page Rank jest obliczany w czasie rzeczywistym."

Anchor Text

Google może użyć danych dotyczących anchor tekstu, żeby wyliczyć Score Rank dokumentu:

- Zmiany w anchor tekście mogą wskazywać na „aktualnienie lub zmianę tematyki” strony/podstrony.
- Anchor text, który przestał być tematycznie związany ze stroną, do której się odnosi może zostać zlokalizowany i zignorowany w razie konieczności. Duże ilości zmiany w dokumencie spowodują sprawdzenie przez wyszukiwarkę powiązanych anchor tekstów i ocenę ich spójności z dokumentem.

- Świeżość anchor tekstu może być wyliczona za pomocą:
- Daty pojawienia się/zmiany anchor tekstu
- Daty pojawienia się/zmiany strony, do której anchor tekst prowadzi
- Daty pojawienia się/zmiany strony, na której znajduje się link
- Google podkreśla, że data pojawienia się/zmiany strony, na której znajduje się link czyni anchor tekst „bardziej odpowiednim i lepszym”

Traffic - Ruch

Google może również ocenić ruch na stronie/podstronie i uwzględnić uzyskane dane przy wyliczaniu Score Rank dokumentu.

- „Duży spadek ruchu na stronie może świadczyć o tym, iż przestała ona być aktualna”
- Google może porównać średni ruch na stronie/podstronie przez ostatnie „j dni” (na przykład j=30) do średniego ruchu na stronie/podstronie przez ostatni rok, żeby sprawdzić czy strona w dalszym ciągu jest odpowiednia dla danego zapytania.
- Google może również użyć sezonowości, żeby ocenić czy konkretna strona/podstrona jest bardziej/mniej odpowiednia dla zapytania podczas określonych okresów w ciągu roku.
- Google może także zmierzyć tzw. "advertising traffic":
- „zakres oraz szybkość z jaką reklamy są prezentowane lub aktualizowane w danym dokumencie w określonym czasie”
- „Jakość reklamodawców”. Google zaznacza, że reklamy takich firm jak Amazon.com będą obciążane większym zaufaniem i oceniane wyżej niż reklamy np. „stron pornograficznych”.
- „The click-through rate” (współczynnik klikalności) uzyskany z ruchu powstałego ze stron na których są reklamy.

User Behavior - Zachowania użytkowników

Google może mierzyć „sumaryczne zachowania użytkowników”. Można tu zaliczyć:

- „Ilość razy jaką dokument jest wybrany spośród wyników wyszukiwania” (CTR w wynikach)
- „Ilość czasu jaką jeden lub więcej użytkowników spędza na danej stronie/podstronie”
- Względna „ilość czasu” poświęcaną na daną stronę/podstronę w prównaniu do średniej ilości czasu.
- Google podaje przykład podstrony z rozkładem godzinowym pływalni, na której użytkownicy zazwyczaj spędzali 30 sekund, a ostatnio nie poświęcają tej podstronie więcej niż „parę sekund”.
- Google twierdzi, że może o twskazywać, że podstrona „zawiera nieaktualny rozkład godzinowy pływalni” i Score Rank takiej strony zostanie obniżony.

Informacje powiązane z domeną

Informacje powiązane z domeną również mogą zostać wykorzystane przez Google przy ocenie dokumentu. Wspomina się o kilku specyficznych rodzajach „informacji związanych z tym w jaki sposób dany dokument jest hostowany w sieci (Internecie, intranecie, itd.)” wliczając w to:

W patencie Google zauważa że wartościowe domeny, są „często opłacane są na kilka lat z góry”, natomiast tzw. 'throwaway domains' rzadko kiedy używane są dłużej niż przez rok”.

Rejestr z DNS może również zostać sprawdzony w celu ustalenia:

- Kto zarejestrował domenę
- Adresy i kontakt do admina oraz obsługi technicznej
- Dane name serwerów
- Stabilność hostowania danych (i firmy hostingowej) vs. duża ilość zmian

Google twierdzi, że może zostać użyta „lista znanych ‘złych’ informacji kontaktowych, name serwerów i/lub adresów IP”, żeby wykryć domeny spamerskie.

Google twierdzi że może wykorzystać również w podobny sposób informacje dotyczące konkretnych name serwerów:

„Dobry name serwer może zawierać mix różnych domen, różnych rejestratorów i posiadać historię hostowania tych domen, natomiast ‘zły’ name serwer będzie hostował głównie pornografię, doorway pages, Made for Adsense lub dużą ilość domen od jednego rejestrującego, może to być też zupełnie nowy name serwer.

Historia rankingu, pozycji strony

Google może zwracać uwagę na historię rankingu (pozycji) danej strony oraz dane z tym związane.

Między innymi:

- Strona której „pozycja skacze przy wielu zapytaniach może oznaczać tematyczny dokument lub próbę spamu”
- „Ilość lub tempo zmiany pozycji dokumentu w rankingu może wpłynąć na przyszłą ocenę tego dokumentu”
- Strony mogą zostać ocenione w zależności od ich pozycji w wynikach wyszukiwania. Najwyższa pozycja zostaje w takim wypadku oceniona najlepiej, natomiast niższe pozycje odpowiednio gorzej.

Google używa równania:

$$[(N+1)-\text{SLOT}]/N$$

Gdzie N=ilość wyszukiwań a SLOT to pozycja w rankingu ocenianej strony

W tym równaniu 1-szy wynik ocenia się na 1.0 natomiast ostatni wynik otrzymuje ocenę bliską 0.

- Google może sprawdzać w szczególności wyniki na „komercyjne zapytania” i dokumenty, które zyskają x% w rankingu mogą zostać oznaczone, lub uzyskany

procent w rankingu może zostać użyty do określenia prawdopodobieństwa ryzyka spamu.

Google może też monitorować:

- „Ilość wejść z wyników wyszukiwania na daną stronę/podstronę w określonym czasie”
- Sezonowość – wahania zachodzące w określonym czasie w roku/miesiącu
- Nagłe zwiększanie się lub zmniejszanie liczby kliknięć
- Inne dane oferowane przez CTR

Wskaźnik zmian w ocenach (score rank) dokumentów w określonym czasie może być mierzony, żeby określić czy szukany termin staje się bardziej konkurencyjny i czy potrzebna jest dodatkowa uwaga.

Google „może monitorować Score Rank dokumentów, żeby wykryć nagłe wzrosty w rankingu”. Takie wydarzenie, według patentu, może oznaczać „albo jakieś aktualne zjawisko np. ‘gorący temat’ albo próbę spamu”

Google może przeciwdziałać próbom spamu w następujący sposób:

„Wykorzystując histerezę (opóźnienie w reakcji na czynnik zewnętrzny) umożliwić nagły wzrost lub spadek Score Rank do określonego (niezbyt dużego) tempa” – histereza w tym wypadku może oznaczać ‘podciąganie’ wyniku przy nagłym obniżeniu wskaźnika wzrostu.

- Ograniczając „maksymalną wartość progową wzrostu w określonym czasie” dla danej strony/podstrony.
- Google „weźmie również pod uwagę wzmianki i linki do strony umieszczane w notkach prasowych, artykułach, na forach i grupach dyskusyjnych itp. zakładając wysoką wiarygodność takich miejsc”

Pewne typy stron (Google w wspomina o „dokumentach rządowych, katalogach stron (np. Yahoo) oraz stronach z relatywnie stabilnym i wysokim Trust Rankiem”) mogą być zwolnione z dodatkowych obserwacji lub nakładanych kar w wypadku nagłych skoków w rankingu.

Google może „również odbierać ewentualne spadki w rankingu jako wskazówka, iż dane dokumenty przestały być popularne lub są nieaktualne”

User Generated Data

Google może również mierzyć różnego rodzaju dane gromadzone w komputerach użytkowników a dotyczące np. ulubionych i często odwiedzanych stron, w tym:

- Lista bookmarków i ulubionych stron w przeglądarce
- Google może pobierać te dane za pomocą takich „asystentów przeglądania” jak Toolbar’y i Desktop Search.
- Dane mogą także pochodzić także bezpośrednio z przeglądarki – Google Chrome ?
- Google może wykorzystać te dane do określenia w jakie strony są wartościowe dla użytkowników

Google chce również dokumentować to, które strony zostają dodane a które usunięte z bookmarków/ulubionych i wykorzystać to przy ocenie strony.

Google może również mierzyć jak często użytkownik wyświetla daną stronę/podstronę i oceniać czy wciąż jest odpowiednia dla danego użytkownika czy jest jedynie pozostałością przeszłych wyświetleń (teraz strona już przestała być popularna lub jest nieaktualna)

„Pliki temp oraz cache powiązane z użytkownikami również mogą być monitorowane” przez Google tak aby zidentyfikować pewne zależności w odwiedzanych stronach i ocenić czy zainteresowanie daną stroną wzrasta czy maleje.

Unique Word, Bigrams, Phrases in Anchor Text

Google buduje również profil tego w jaki sposób pojawia się anchor tekst do danej strony/podstrony, żeby wykryć spam. Zaznaczają, że „web graph” (wykres, wektor strony, wektor Page Rank ?) rozwijający się naturalnie zazwyczaj jest wypadkową indywidualnych decyzji, natomiast wektor strony generowany syntetycznie (często związany z próbami spamu) opierają się na skoordynowanych działaniach. Różnice w obu schematach rozwoju mogą zostać zmierzone i użyte aby blokować spam.

Google zaznacza, że duże przyrosty „anchor words/bigrams/phrases” są głównym celem pomiarów. Według nich spam w tym kontekście to „dodawanie dużej liczby identycznych anchorów z wielu dokumentów”.

Linkage of Independent Peers

Google może użyć również danych dotyczących linków od „niezależnych peerów (np. dokumentów nie powiązanych” w celu sprawdzenia spamu. Twierdzą, iż „nagły wzrost peerów niezależnych... z dużą ilością linków... może wskazywać potencjalnie syntetyczny web graph, który jest wskazaniem próby spamu”. Google twierdzi, że „podejrzenie spamu może zostać wzmocnione jeśli wzrost powiązany jest z anchor tekstem, który jest nadzwyczajnie zgodny lub niezgodny” i mogą obniżyć znaczenie takich linków za pomocą ‘stałej ilości’ (fixed amount) lub ‘czynnika mnożącego’ (multiplicative factor) – co może być dodatkową karą za samo posiadanie takich linków.

Tematyka dokumentu

Określenie tematyki danej strony może zostać przeprowadzone przy użyciu:

- Kategoryzacji
- Analizy URL
- Analizy zawartości
- Clusteringu
- Podsumowania
- Zestawu unikalnych, rzadko spotykanych słów

Celem jest „monitorowanie tematu/-ów danego dokumentu w określonym czasie i użycie zebranych danych przy ocenie”.

Google zaznacza, że „nagły wzrost w ilości tematów może wskazywać spam” oraz, że znaczące zmiany w temacie mogą wiązać się ze „zmianą właściciela strony i poprzednie cechy dokumentu, takie jak Score Rank, anchor tekst itp. nie są już wiarygodne.” Google podkreśla, że „jeśli wykryją jedną lub więcej takich sytuacji, zredukowany zostanie Score Rank takich dokumentów a wartość linków, anchor tekstu oraz innych danych zostanie obniżona”.

Jeśli w plikach pomocy Google czytamy że za pozycję w Google odpowiadać może nawet 200 czynników... to analizując powyższy patent już możemy doszukać się co najmniej 50-60 tego typu zmiennych decydujących o Score Rank naszej strony...

Chciałbym także zwrócić uwagę na nowy termin, który stosujemy w tym artykule - **Score Rank** - czyli innymi słowy Real Page Rank... Myślę że używanie w tym kontekście słowa "Page Rank" za mocno piętnuje i kojarzy się z zielonym paskiem Google. Tymczasem Score Rank odnosi się do oceny dokumentu i w efekcie do pozycji w wynikach organicznych i naszym zdaniem Score Rank doskonale opisuje ideę dbania o dobre pozycje. Suma 200 czynników algorytmu Google (czyli m.in. także Page Rank) daje nam efekt w postaci Score Rank danego dokumentu na daną frazę.

Druga połowa tekstu jest głównie tłumaczeniem z SEOmoz.org... zachęcam także do zapoznania się z [oryginalną wersją patentu Google!](#)
Dziękuję Jackowi Kubiakowi za pomoc w tłumaczeniu.