



CTI jelentés

A keresőmotorok működése





Tartalomjegyzék

Bevezetés	4
A keresőmotorok története	6
Mi az a keresőmotor?	9
• Algoritmusok	9
• Feltérképezés	18
• Indexelés	21
Miért éri meg a keresőmotornak?	27
• Hirdetések	27
• Online kereskedelem	29
• Szolgáltatások	29

Keresőmotorokkal elkövetett csalások	30
• SEO poisoning	30
• Blackhat SEO	32
- <i>Keyword Stuffing</i>	32
- <i>Cloaking</i>	33
- <i>A keresési rangsor manipulálása</i>	33
- <i>Private Link Networks</i>	34
- <i>Article Spinning</i>	34
- <i>Sneaky Redirects</i>	36
- <i>SEO spam vagy spamdexing</i>	36
- <i>SEO scam</i>	37
• Egyéb technikák	38
- <i>Injection</i>	38
- <i>Siteurl hackek</i>	38
- <i>HTML injekciók</i>	39
- <i>Adatbázis injektálások</i>	39
- <i>Tetszőleges fájlbefecskendezések</i>	40
Összegzés	43

Bevezetés

A legismertebb keresőmotor a Google, de számos más fajta is létezik. A világon körülbelül 30 olyan kereső létezik, amely kizárólag általános internetes keresésre szolgál, azonban ha hozzávesszük a speciális funkciókat ellátó és a különböző országok sajátosságaihoz tartozó réskeresőmotorokat is, akkor **a számuk világszerte több mint 1500-ra nő.**

A legnépszerűbbek közé tartozik a Bing, a Yahoo!, a DuckDuckGo és a Yandex. Ezek a keresőmotorok különböző módokon működhetnek. A legáltalánosabb technika amit használnak, az a **weboldalak indexelése, a kulcsszavak elemzése és a rangsorolás**, így nyújtják a felhasználók számára a legrelevánsabb találatokat.



A népszerűsége okán, cikkünkben ahol szükséges a Google keresőn keresztül mutatjuk be a funkciókat.

A keresőmotorok az interneten található információk hatalmas mennyiségét teszik kereshetővé, és lehetővé teszik a felhasználóknak, hogy gyorsan és hatékonyan megtalálják a keresett tartalmat, weboldalt, képeket, videókat vagy cikkeket.

Legtöbbször Intézetünk javaslatai között is szerepel (ahogy korábbi [anyagunkban](#) is említettük), hogy **inkább keressünk rá egy intézmény weboldalára, mint sem adathalászat áldozatává váljunk az által, hogy rákattintunk egy hamis/hamisított hivatkozásra.**

De hogyan is működnek ezek a keresők és milyen buktatói lehetnek?



A keresőmotorok története



DAVID FILO és **JERRY YANG** létrehozta a **Yahoo! Search**-ot. Ez kezdetben egy gyűjtemény volt, amely **tartalmazta a weboldalak listáját**. Az **URL-ekhez ember alkotta leírásokat is mellékeltek**. Az oldalak tulajdonosai **információs oldalakat adhattak hozzá**, viszont a **kereskedelmi oldalaknak ezért 300 dollárnyi éves díjat kellett fizetniük**.

A **Lycos** keresőmotor szintén 1994-ben indult, amely **augusztusig több mint 394 000 dokumentumot katalogizált**. 1995 év elejére ezek száma meghaladta a 1,5 milliót.

1994

Ugyanabban az évben elindult a **Web Crawler** is. Ez volt **első olyan keresőmotort, amely egész oldalakat indexelt**.



TIM BERNERS-LEE, a CERN egyik dolgozója hozta létre a világhálót.

Ő fejlesztette ki az **első webszervert**, az **első webböngészőt** és **egy dokumentumformázási protokollt**, a **Hypertext Markup Language (HTML)** leíró nyelvet.

1989

1993 FEBRUÁR

HAT STANFORDI DIÁK alkotta meg az **Excite**-ot, amelynek alap gondolata, hogy a **keresés hatékonyabbá tételéhez a szókapcsolatok statisztikai elemzését kellene használni**. Az év közepén adták ki keresőszoftverük másolatait.

MARTIN KOSTER a Wandererre válaszul **létrehozta a WWW Archie-szerű indexelését, vagyis az ALIWEB-et**. Az ALIWEB **a metainformációkat térképezte fel, és lehetővé tette a felhasználók számára, hogy saját leírással ellátott oldalakat küldjenek be**, amelyeket indexelni szeretnének. Az ALIWEB-nek **nem volt szüksége botra az adatgyűjtéshez, és nem használt nagy sávszélességet**.

1996

A Berkeley Egyetem professzora **ERIC BREWER**, és **PAUL GAUTHIER** végzős hallgató megalapította az **Inktomi**-t. A cég volt a **fizetett hirdetési modell pay-per-click (PPC) úttörője**. A Yahoo! 2003-ban 235 millió dollárért megvásárolta.

LARRY PAGE és **SERGEY BRIN** a Stanford Egyetemen megalkotta a **Google**-t, ami a hivatkozás népszerűsége vizsgálató **PageRank*** algoritmuson alapul. A korabeli keresőkhöz képest nagyon fejlett sorrendezésének köszönhetően gyorsan kiszorította a vetélytársait, és **2001 körül egyeduralmukodóvá vált**.

1993 OKTÓBER



1993 JÚNIUS

MATTHEW GRAY bemutatta a **World Wide Web Wanderert**, azaz a **Wandexet**. Először a **web növekedését akarta mérni, és az aktív webszerverek számlálására hozta létre a WWW botot**, amelyet **később továbbfejlesztett, hogy tényleges URL-címeket rögzítsen**.

Ez volt az **első szoftver, amely már a WWW-on keresztül is képes volt keresni**.

2011

A strukturáltabb internet létrehozása érdekében a **GOOGLE**, a **MICROSOFT**, a **YAHOO** és a **YANDEX** létrehozta a **Schema.org**-ot. A Schema.org célja a **strukturált adatok sémáinak létrehozása, karbantartása és népszerűsítése az interneten, weboldalakon és e-mail üzenetekben**. A közös szókincs megkönnyíti a webmesterek és a fejlesztők számára, hogy milyen séma mellett döntsenek és ezáltal képesek a hatékonyságukat maximalizálni.



2015

A **GOOGLE** kiadta a **Mobilegeddon** algoritmust, ezzel rákényszerítve a weboldalakat a mobilbarát weboldalak hozzáadására. Ezéven a Bing is bevezette mobilbarát algoritmusát.

A **MICROSOFT** elindította az **MSN/Live Search** keresőmotorját, amely **2009-ben megkapta a Bing nevet**. Ugyan ebben az évben a Google **HTTPS Everywhere** kampánya révén a **webhelyek biztonsági szintje is egy rangsorolási tényezővé vált**.

2004

*A PageRank segítségével a Google elemezni tudja a különböző oldalak közötti kapcsolatokat, és ennek segítségével sokkal relevánsabb találatokat tudott visszaadni, mint vetélytársai, amelyek az oldalak sorrendezését kizárólag azok tartalma alapján határozták meg.

A Mobilegeddon megjelenéséig a világ legnagyobb webhelyei közül is jópáran úgy gondolták, hogy teljesen elfogadható, ha a mobilfelhasználókat is asztali tartalmakkal szolgálják ki. 2013-ban az összes webes forgalomnak mindössze 16%-a történt mobil eszközön keresztül.

Először 2017-ben fordult elő, hogy mobil eszközökön **nagyobb adatforgalmat bonyolítottak le** a felhasználók, mint az asztali számítógépeken.



TUJTAD?

A mobil eszközök nem csak a mobiltelefonokra, hanem a **laptopokra, a táblagépekre, sőt még az okosórákra is vonatkozik**. Mindegyik eszköz más képernyőmérettel és képfelbontással rendelkezik. Ha egy webhely **nem rendelkezik reszponzív webdesignnal**, akkor **néhány eszközön jól fog megjelenni, míg másokon teljesen szét tud csúszni** a megjelenítendő tartalom.

A reszponzív webdesign olyan megközelítést jelent, **amely lehetővé teszi, hogy a weboldal tartalma "reagáljon" a használt eszközök képernyő- és ablakméretére, és alkalmazkodjon ahhoz**. Biztosítja az optimális böngészést és felhasználói élményt függetlenül attól, hogy épp mobilon, tableten vagy asztali gépen végezzük azt.



Mi az a keresőmotor?

A keresőmotorok lehetővé teszik a felhasználók számára, hogy az interneten kulcsszavak segítségével keressenek tartalmakat. Amikor beírunk a keresőmotorba egy kifejezést, **egy találati oldal (SERP) jelenik meg, amely az oldalakat relevanciájuk szerint rangsorolja. Ennek módja keresőmotoronként eltérő.**

Ezt különböző algoritmusokkal teszik, amelyeket gyakran változtatnak a felhasználói élmény javítása és a visszaélések elkerülésének érdekében. Mindezt azért, hogy a felhasználóknak a legjobb választ adják a lekérdezésükre, a legjobb minőségű és legrelevánsabb oldalakat helyezik előtérbe a keresők.

Algoritmusok

A leghíresebb Google algoritmusok kronológiai sorrendben:

→ 1996: PageRank

A Sergey Brin és Larry Page által létrehozott korai projekt. Ez az algoritmus erős alapot teremtett az internet automatikus indexeléséhez, de még sebezhető volt a black hat taktikákkal szemben.

→ 2003: Google Florida

Ez volt az első algoritmus, ami büntetést szabott ki olyan vétségekért, mint a keyword stuffing.





→ **2004: TrustRank**

Az algoritmus egy lépéssel túlmutatott a Floridán, mivel a SPAM és a black hat technikák kiszűrésére szolgált a többi keresőmotor találatai közül. Ez az algoritmus segít azonosítani, hogy a domainek mennyire megbízhatóak, így a felhasználók minőségi eredményeket kapnak.

→ **2009: Google Caffeine** 

Gyorsította a keresést. Ehhez később egy olyan funkció is megjelent, amivel valós idejű kereséseket folytathattunk le, és figyelembe vették a tweeteket és a legfrissebb híreket is.


→ **2011: Google Panda** 

Egy olyan algoritmusként debütált, ami a tartalom farmokra specializálódott. Ezek olyan oldalak voltak, amelyeken nagy mennyiségű, de alacsony minőségű tartalom jelent meg. Gyakran frissítették azzal a kizárólagos céllal, hogy a keresőmotorok értékelésénél jól szerepeljenek. Emellett jellemző volt rájuk, hogy a tartalom és a hirdetések aránya nagyjából hasonló volt, amit a Panda könnyedén megtalált.

→ **Google Freshness**

A Google Freshness algoritmus pontosan azt teszi, amit várnánk; a friss, releváns tartalmat helyezi előtérbe.

Ez az algoritmus egyébként a Caffeine algoritmus továbbfejlesztése volt.

→ **2012: Google Penguin** 

Azokat az oldalakat célozta be, amelyek finoman alkalmaztak kevésbé elfogadott optimalizálási taktikákat.

Például olyan oldalakat, amelyek főleg tájékoztató tartalommal rendelkeztek, de elhintve sok olyan hiperlink is megtalálható volt, amelyek spam jellegűek és nem kapcsolódtak szorosan a tematikához.

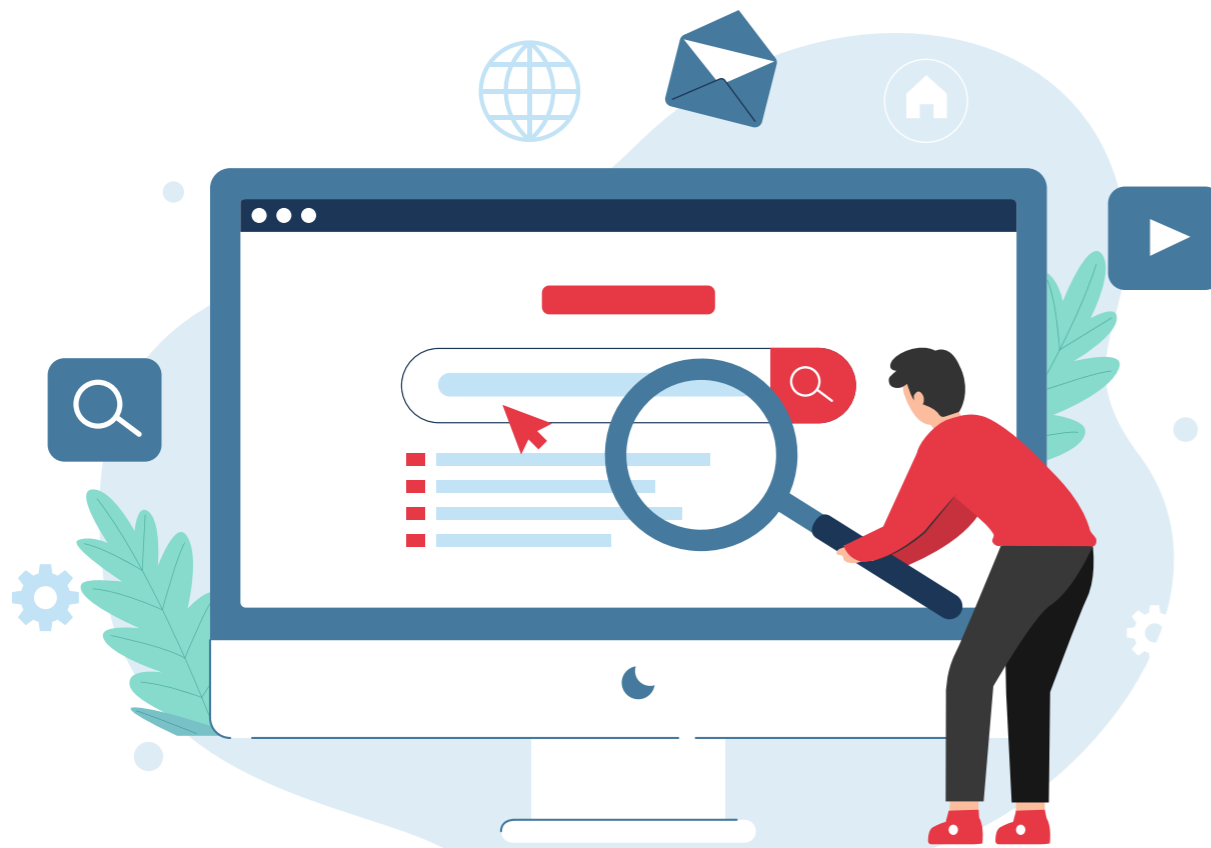
→ **2013: Google Hummingbird** 

Bevezetésével már a felhasználók szándékát igyekezte kitalálni a Google, és nem csak azt nézte, hogy milyen szavakat írnak be a keresőbe. Megfigyelték, hogy az emberek már nem kulcsszavakra, hanem komplett kérdésekre is keresnek. Olyan tényezőket is figyelembe vett a találati lista kialakításánál, hogy honnan és milyen eszközzel internetezett a felhasználó, melyek az érdeklődési körei, korábban milyen oldalakat látogatott meg, mire keres általában.



2014: Google Pigeon

Az elsődleges cél a térképes keresések javítása volt. Emellé olyan funkciókat kapott, mint például helyesírás ellenőrzés és szinonimák felvetése.



Payday Loan

Azokra a találatokra koncentrált, amelyek nagyobb valószínűséggel hoznak spam jellegű eredményeket. A Google úgy igazította a rangsorolás módját, hogy ezek az oldalak ne érjenek el kiemelt helyezést.

2015: Google Mobilegeddon

Célja az volt, hogy a mobilszközökön indított kereséseknél, a mobilbarát oldalak előnyhöz jussanak. Azok az internetes oldalak, amelyek nem támogatták az okoseszközökről történő elérést, hátrébb sorolódtak.

Google Rankbrain

A fő feladata az volt, hogy megértse, mit keresünk, akár idézőjelek nélkül is.

2016: Google Possum

A helyi vállalkozásoknak lehetőséget kínált arra, hogy kapcsolatba lépjenek lokális közönségükkel. Ha rákeresünk a "kerékpár javító műhely" kifejezésre, akkor nem a világ összes ilyen vállalkozását tartalmazó univerzális keresés eredményeit adja vissza, hanem csak a saját régiónk személyre szabott műhelyei kerülnek listázásra.



2017: Google Fred

Az algoritmus tovább igyekezett büntetni a rossz weboldalakat, amelyek túl sok hirdetést és kevesebb minőségi tartalmat kínáltak.



2018: Google Medic

Ez az algoritmus segített abban, hogy az orvosi tekintéllyel rendelkező webhelyeket előnyben részesítse az orvosi tekintéllyel nem rendelkező webhelyekkel szemben.

2019: Google Bert

Egy óriási előrelépés történt a felhasználók keresési szándékának megértésében és egyben a jobb minőségű találati listák elkészítésében.



Search

2020: Passage Ranking

Az Egyesült Államokban indult, és az angol nyelvű lekérdezéseket segítette elő. Most már a weboldal egyes szakaszait értékelte az algoritmus, hogy jobban megértse, mennyire releváns egy oldal a keresés szempontjából. Továbbra is a teljes oldalakra vonatkozó információk alapján rangsorolt.

2021: MUM

A MUM a „Multitask United Model” rövidítése, amely név utal az új algoritmus erejére: egyszerre több feladatot is képes kezelni. Több mint 75 nyelven képes olvasni, megérteni és tanulni, többféle forrásból, most már videóból és hangból is.

Page Experience

2021-ig fokozatosan került bevezetésre. Egy új rangsorolást vezetett be, amely „a valós felhasználói élményt mérő metrikákat tartalmaz, a betöltési teljesítmény, az interaktivitás és a vizuális stabilitás tekintetében”.

 **2022: Hasznos tartalom**

A keresési eredményekben megjelenő tartalom minőségére helyezte a hangsúlyt. A lényeg az „emberközpontú” tartalmak jutalmazása - az olyan tartalmaké, amelyek valóban választ adnak a felhasználók kérdéseire és kielégítő élményt nyújtanak.

Ez a frissítés különösen a sok alacsony minőségű tartalommal rendelkező oldalakat érintette, amelyek ennek következtében rosszabbul szerepelhettek a keresőben.

 **2023: “Nyitottság”**

Helyi keresési frissítés: az, hogy egy üzlet jelenleg nyitva van-e, erősebb pontszám lett a nem navigációs lekérdezéseknél.

 **Témahatósági rendszer**

A rendszer célja, hogy segítse a releváns, szakértő és hozzáértő tartalmak felszínre kerülését a Google keresőben és a Google Newsban. A rendszer számos jelet vizsgál, hogy megértse egy adott témával vagy helyszínnel kapcsolatos publikáció szakértelmét.

 **2024: Visszaélés a webhely hírnevével**

Az elsősorban rangsorolási céllal és a webhely tulajdonosának szoros felügyelete nélkül előállított, harmadik féltől származó tartalmak spamnek minősülnek. Egyes SEO-k ezt a gyakorlatot „parazita SEO”-nak nevezik.

 **Spam frissítések**

A keresési rangsorok növelését célzó tartalom előállítása (akár automatizálással, akár emberekkel, akár ezek kombinációjával) spam kategóriájába került besorolásra. A lejárt domainek megvásárlása és újrahasznosítása az alacsony minőségű tartalom keresési rangsorának növelésének szándékával, spamnek minősül.

A legtöbb keresőmotor a hivatalos weboldalán ad tippet, hogy milyen módszerekkel érhetünk el jobb rangsorolást, valamint rengeteg leírás érhető el a neten.

Feltérképezés

A Google első indexe 1998-ban 26 millió oldalt tartalmazott. 2000-re elérte az egymilliárdos, 2008-ra az “1 trillió egyedi URL egyszerre a weben” mérföldkövet.

Folyamatosan bővítik és frissítik a weboldalakról, képekről, könyvekből, videókból, tényekből és sok más tartalomtól származó adatokkal.

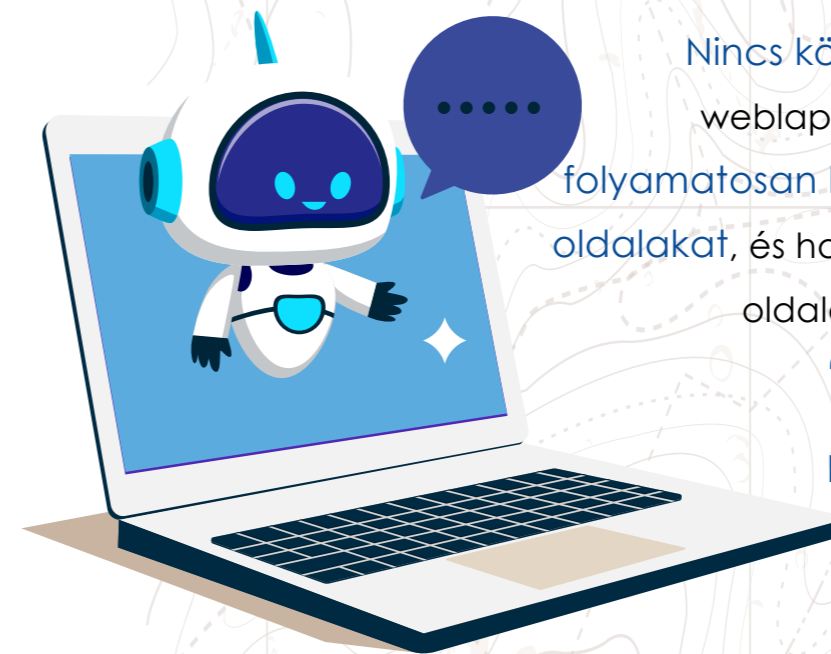
A Google teljesen automatizált, keresőrobotként ismert szoftvereket használ, amelyek rendszeresen vizsgálják az internetet, hogy megtalálják az indexünkbe felvehető oldalakat. A találatokban szereplő oldalak túlnyomó többségét nem kézzel küldték be a keresőbe, hanem automatikusan találják meg és adják hozzá a keresőrobotok.

A keresés általában három szakaszban történik meg és nem minden oldal jut át minden szakaszon.



A három szakasz az alábbiak szerint alakul:

- ▶ **Lekérdezés:** A keresőrobot az interneten talált oldalakról, automatizált programokkal szövegeket, képeket és videókat tölt le.
- ▶ **Indexelés:** Elemzik az oldalon található szöveget, képeket és videofájlokat, és az információkat az indexben tárolja, amely egy nagy adatbázis.
- ▶ **Keresési eredmények kiszolgálása:** Amikor egy felhasználó keresést indít, a weboldal ennek megfelelő információkat küld vissza.



Nincs központi nyilvántartás az összes weblapról, ezért a keresőrobotoknak folyamatosan keresniük kell az új és frissített oldalakat, és hozzá kell adnia azokat az ismert oldalak listájához. Ezt a folyamatot “URL-felfedezésnek” nevezik.

Egyes oldalak azért ismertek, mert egy keresőrobot már meglátogatta őket.

Más oldalakat akkor fedeznek fel, amikor egy ismert oldalról egy új oldalra mutató linket követ. Egy harmadik lehetőség, amikor elküldjük az oldalak listáját (a webhelytérképet) a Google számára, hogy térképezze fel azokat.

A keresőrobotok folyamatosan, több milliárd oldal feltérképezését végzik egyszerre a weben, és miután felfedezik egy oldal URL-címét, meglátogatja (vagy "feltérképezi") az oldalt, hogy megtudja, mi található rajta.

A keresőrobot egy algoritmikus folyamat segítségével határozza meg, hogy mely oldalakat és milyen gyakran kell feltérképeznie. Ezek általában úgy vannak programozva, hogy az oldalak lassulásának elkerülése érdekében igyekeznek nem túl gyorsan feltérképezni azokat. Ez a mechanizmus a webhely válaszain alapul (például a HTTP 500-as hibakódok).

A keresőrobotok azonban nem kutatják át az összes felfedezett oldalt, mivel egyes oldalakon a webhely tulajdonosa letilthatja a feltérképezést, más oldalak pedig bejelentkezés nélkül nem érhetőek el.



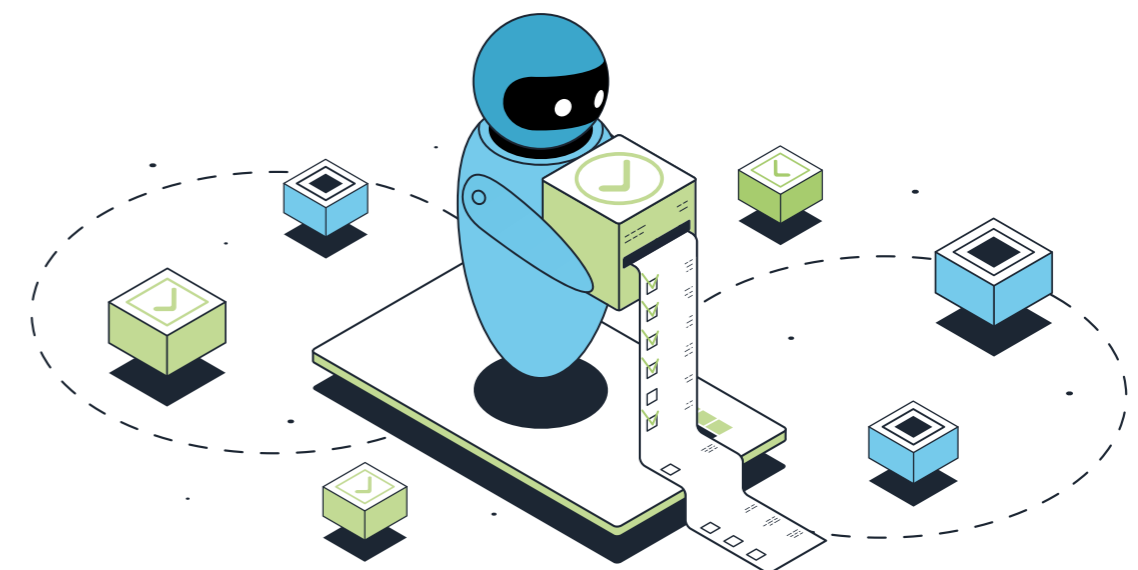
A feltérképezés során a keresőrobot letölti a HTML-t és kigyűjti belőle a linkeket. Ha JavaScript-et talál, letölti az oldal CSS és JavaScript fájljait, majd a Web Rendering Service segítségével lefordítja majd megjeleníti, indexeli az oldalt, és kigyűjti belőle a hivatkozásokat. Mindkét esetben ez után elküldi azokat a Google indexelő motorjának.

Indexelés

Miután egy oldalt feltérképeztek, a keresőmotor megpróbálja megérteni, hogy miről szól. Ezt a szakaszt indexelésnek nevezik, és magában foglalja a szöveges tartalom, valamint a kulcsfontosságú tartalmi címkék és attribútumok, képek, videók stb. feldolgozását és elemzését.

Az indexelési folyamat során a kereső megpróbálja meghatározni, hogy egy weblap egy másik internetes oldal duplikátuma vagy kanonikus-e. Kanonikus az az oldal, amely megjelenhet a keresési eredményekben.

A kanonikus kiválasztásához először csoportosítja (más néven klaszterezi) az interneten talált, hasonló tartalmú oldalakat, majd kiválasztja a csoport legrepresentatívabb példányát. A csoportban lévő többi oldal alternatív változatok, amelyeket különböző kontextusokban lehet kiszolgálni, például ha a felhasználó mobil eszközről vagy ha egy nagyon konkrét oldalt keres az adott klaszterből.



A keresők jeleket is gyűjtenek a kanonikus oldalakról és annak tartalmairól. Néhány jel közé tartozik az oldal nyelve, az ország, amelyre a tartalom lokálisan vonatkozik, valamint az oldal használhatósága.

Az összegyűjtött információkat egy nagy adatbázisban tárolják, amelyet több ezer szerveren tárolnak. Az indexelés nem garantált; nem minden olyan oldal kerül indexelésre, amelyet a keresőrobotok feldolgoznak.

Az indexelés az oldal tartalmától és metaadataitól is függ.

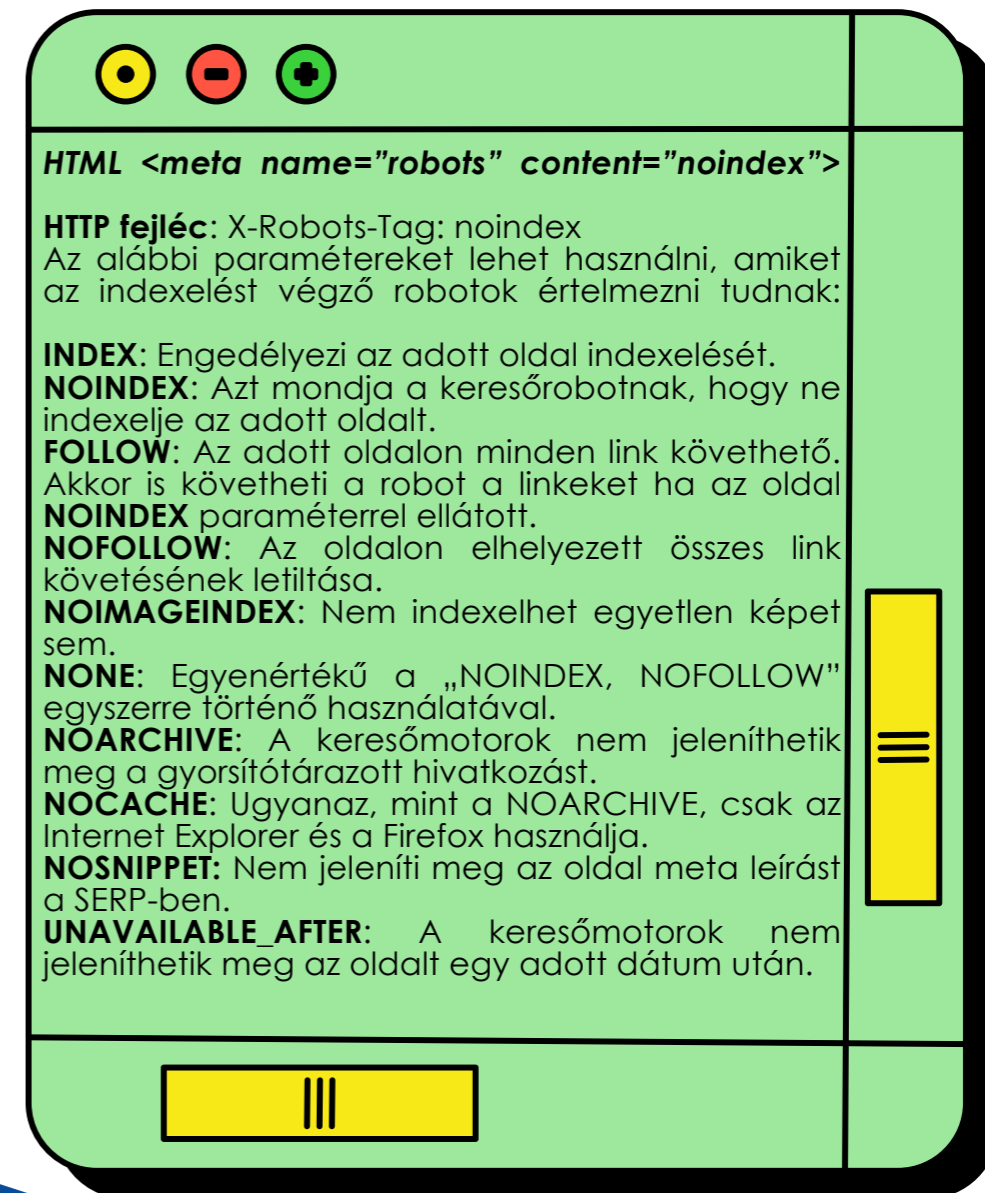
Metaadat: Az oldal HTML kódjának részeként vagy különálló fájlban tárolt információk, mint például a weboldal címe, szerzője, leírása, kulcsszavai stb.



TUJTAD?

Néhány gyakori indexelési probléma lehet:

- ▶ **Az oldal tartalmának minősége alacsony.** A Google az oldal tartalmának minőségét az információ hasznossága, a hitelessége és az átfogó jellege alapján értékeli.
- ▶ A robotok meta szabályai tiltják az indexelést.
- ▶ A HTML oldalak forráskódjába vagy egy HTTP fejlécbe illesztett adatok:



- ▶ A weboldal kialakítása megnehezítheti az indexelést.
- ▶ A keresési eredmények kiszolgálása.

Amikor egy felhasználó beír egy keresést, a keresőmotorok átnézik az indexet a megfelelő oldalak után, és **azokat az eredményeket adják vissza, amelyekről úgy gondolják, hogy a legjobb minőségűek és a leginkább relevánsak** a felhasználó keresése szempontjából.

A relevanciát **több száz tényező határozza meg**, amelyek között olyan információk is szerepelhetnek, mint a felhasználó tartózkodási helye, nyelve és eszköze (pl. asztali számítógép vagy telefon). Például a "kerékpárjavító műhelyek" keresés **más eredményeket mutatna** egy párizsi és egy hongkongi felhasználónak.

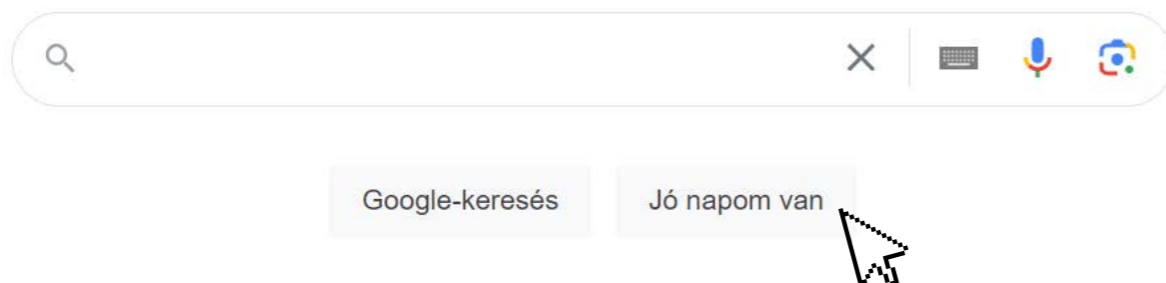


Az alábbiakban található két kép a fentebb említett keresési eredményeket mutatja be. Bal oldalon a párizsi, jobb oldalon pedig a hongkongi eredmények találhatóak:

The left screenshot shows Google search results for "bike repair shop" in Paris, France. The search bar contains "bike repair shop". Below the search bar, there are filters for "Most nyitva van", "Legjobb értékelésű", "Párizs", and "Vincennes". The results section is titled "Úti célokat tartalmazó webhelyek" and lists several websites: "THE BEST 10 Bike Repair/Maintenance in PARIS, FRANCE - Last Updated April...", "Bike Repair Center", "Bike Repair Center - Vente et réparation de vélos et cycles, 246 av Ferrayonnes...", and "Great bike repair shop - Review of Cool bike, Bordeaux, France". Below this, there is a "Vállalkozások" section with a map of Paris and a list of businesses: "MebVelo", "Jour de Vélo Bastille", "Les Cyclistes Branchés", "France Velo Turismo", "cyclofix.co.uk", and "Le Bonbon | PARIS".

The right screenshot shows Google search results for "bike repair shop" in Hong Kong, China. The search bar contains "bike repair shop". Below the search bar, there are filters for "Most nyitva van", "Maximális távolság: 5 mf", and "Legjobb értékelésű". The results section is titled "Helyek" and lists several businesses: "Chat Kee Bicycle Co.", "青公 Bike Station (北角店)", "Decathlon HK", "Cycling Workshop | Maintenance & Repair", "Sky Blue Bikes", "Services - Maintenance", "BullBike", and "Maintenance". Below this, there is a "Mások ezeket a kérdéseket is felteszik" section with several questions: "Is it worth fixing a bike?", "What is called bike mechanic?", "What is the phone number for Velofix Ottawa?", and "Is cycling popular in Hong Kong?".

A felhasználó lekérdezése alapján a keresési találati oldalon megjelenő keresési funkciók is változnak. Például a “kerékpárjavító műhelyek” keresés elsősorban helyi találatokat fog mutatni, és nem képi találatokat, azonban a “modern kerékpár” keresés esetén valószínűbb, hogy képi találatok jelennek meg, de helyi találatok nem.

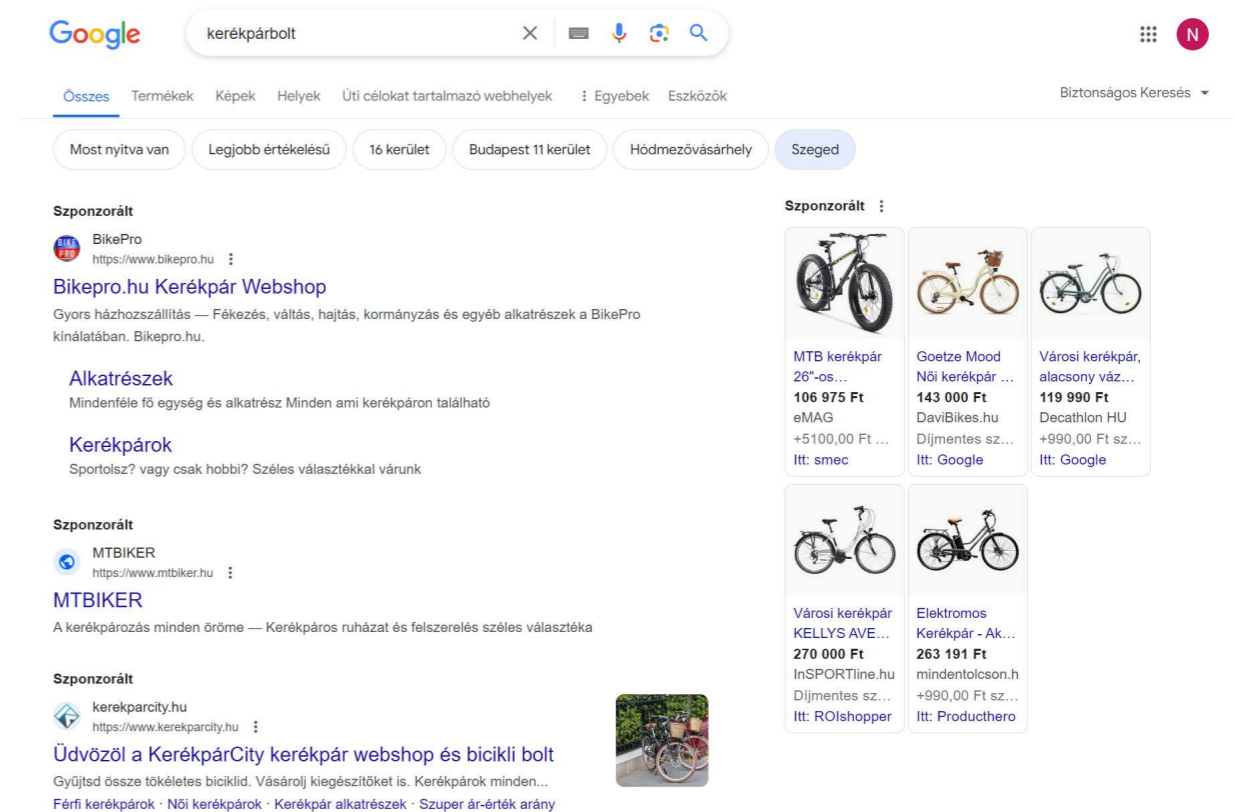
A Google kezdőlapján található egy “Jó napom van” feliratú gomb. Ez a funkció lehetővé teszi a felhasználók számára, hogy ha beírják a keresési lekérdezésüket és rákattintanak a gombra, a találati oldalt kihagyva, közvetlenül az első találatra kerüljenek. Ha a keresőmezőt üresen hagyjuk, a gombra kattintva megnyílik a Google Doodles archívuma.

Miért éri meg a keresőmotornak?

A keresők használata a felhasználóknak ingyenes, így felmerülhet az a kérdés, hogy miért éri meg mindez az üzemeltetőjének.

Hirdetések

A keresés során olyan hirdetések is meg fognak jelenni, amely a termékkel, szolgáltatásokkal kapcsolatos, vagy hasonló tartalmú, mint a látogatni kívánt webhely. Ezek a keresés elején “Szponzorált” címkével jelennek meg. A Google partnerek milliárdnyi weboldalán Google Hirdetések jelennek meg.



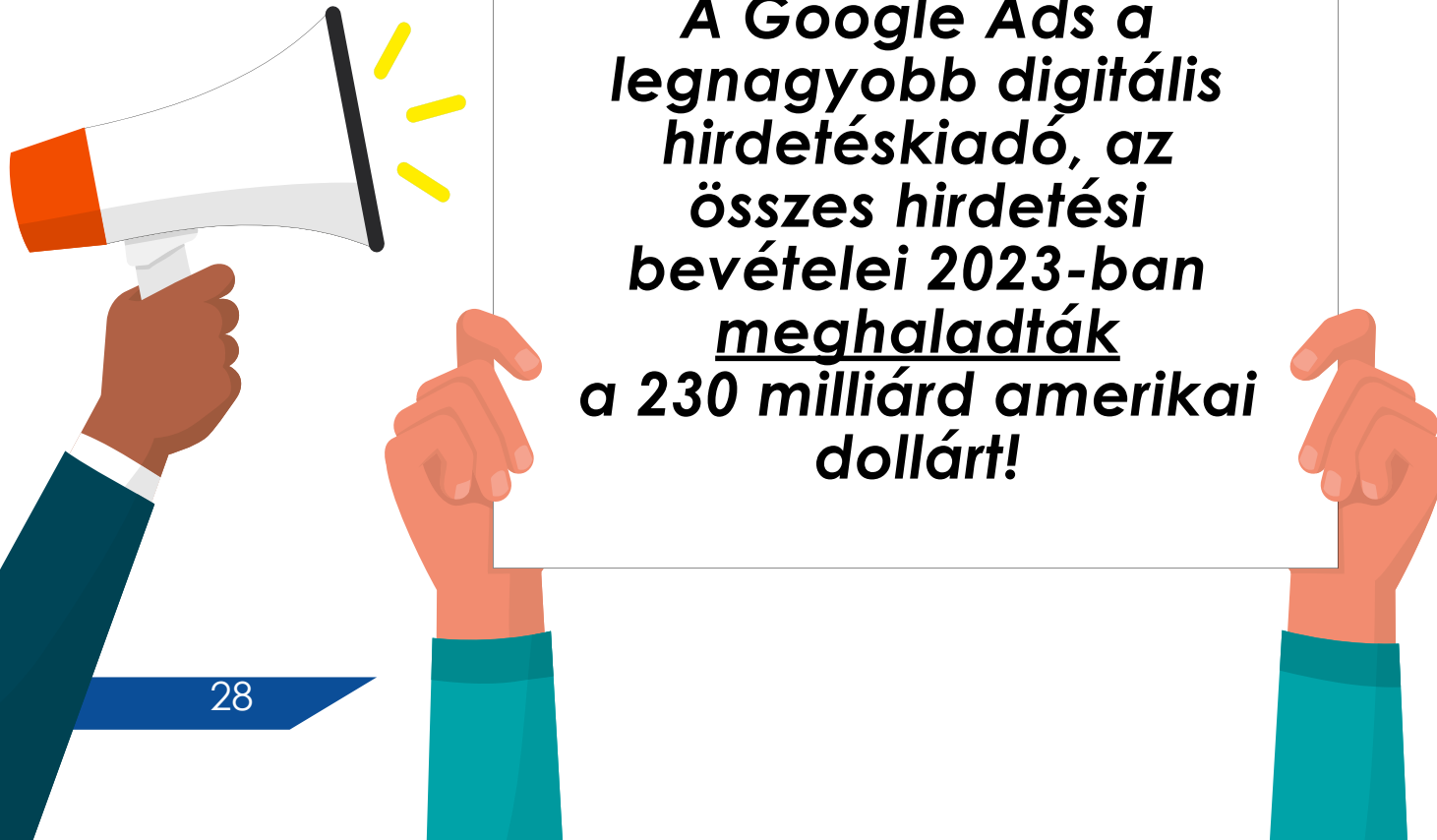
1. ábra

“Kerékpárbolt” Google keresési eredmények, szponzorált linkekkel

Mindez egy villámgyors hirdetésaukció segítségével történik, amelyre minden alkalommal sor kerül, amikor valaki keresést végez a Google rendszerében, illetve olyan webhelyre látogat, ahol hirdetések jelennek meg.

A Google Ads az aukcióban részt vevő **összes hirdetésnek kiszámítja a pontszámát, azaz a hirdetés rangsor szerinti helyét**. Ez a rangsor pontszám határozza meg a pozíciót, illetve azt, hogy megjeleníthető-e egyáltalán a hirdetés.

Ha kattintásonkénti költség (CPC) rendszerű ajánlattételt használ a hirdető, akkor **ebben az esetben csak kell fizetnie, ha valaki a hirdetésre kattintva felkeresi a webhelyet**. A Google Ads rendszerben **beállíthatjuk, hogy maximum mennyit vagyunk hajlandóak fizetni a hirdetésre leadott kattintásokért** (ezt az összeget maximális kattintásonkénti költségajánlatnak nevezzük).



A Google Ads a legnagyobb digitális hirdetéskiadó, az összes hirdetési bevételei 2023-ban meghaladták a 230 milliárd amerikai dollárt!

Online kereskedelem

A keresőmotorok találatai között **megjelenő márkák különféle termékei is fizetett pozíciókban jelennek meg**. Amikor a felhasználók rákattintanak ezekre a termékekre, a keresőmotor gyártója is részesedik egy adott százalékban a bevételből.



A keresőmotorok gyártója számos más online szolgáltatást is kínál, amelyek bár ingyenesek, de **a felhasználók extra funkciókat is feloldhatnak bennük egy kisebb összegért cserébe**. Például e-mail keretrendszert, térképeket, RSS feedet és akár autó navigációt is.

A Google Drive például 15 GB ingyenes tárhelyet kínál, de havi néhány ezer forintért több száz GB-ra bővíthető.

A másik jó példa a sok ingyenes alkalmazással rendelkező Google Play vagy Apple App Store. Az ezekben **megjelenő hirdetések után szintén részesedik** a bolt. Ezek az alkalmazásboltok a fizetős alkalmazások árából is részesednek.

Keresőmotorokkal elkövetett csalások

A keresőmotorok térhódítása következménye, hogy [itt is megjelentek a különféle csalási kísérletek](#). Ha keresőrobotok károsnak ítélik a tartalmat, egy vizsgálat után feketelistára teszik az oldalt. Ezért ha találkozunk ilyen oldallal, érdemes jelentenünk.

SEO poisoning

A SEO poisoning, más néven keresőmérgezés, a rosszindulatú reklámozás (malvertising) egy olyan típusa, amelyben a kiberbűnözők rosszindulatú webhelyeket hoznak létre, majd keresőoptimalizálási (SEO) technikákat használnak arra, hogy a webhelyek linkjei az első találatok között, gyakran hirdetések formájában a keresési eredmények tetején jelenjenek meg.

A találatokban való kiemelkedő szerepük miatt a felhasználók gyakran azt feltételezik, hogy a linkek legitím és biztonságos webhelyekhez kapcsolódnak.

[Security Software - Free Software Downloads and Software Reviews ...](#)
download security center. Your source for antispayware and security downloads ... Software maker will release its Microsoft Security Essentials "in the coming ... Vista and Windows Server 2008, but not the final version of Windows 7. ...
[download.cnet.com/windows/security-software/ - Similar](#)

[Linda Chong's Blog : Download a free copy of Windows Security ...](#)
30 Sep 2009 ... Microsoft Security Essentials is officially released in 8 languages and 19 countries around the world. You can download Microsoft Security ...
[blogs.msdn.com/.../download-a-free-copy-of-windows-security-essentials-to-protect-your-home-pc-and-laptop-today.aspx - 9 hours ago - Similar](#)

[Microsoft Security Essentials Download](#)
29 Sep 2009 ... Free Microsoft Security Essentials available for download Microsoft has good reason to ensure Windows PCs are secure and malware-free. ...
[www. security-essentials-download - 16 hours ago - Similar](#)

2. ábra
SEO poisoning példa

A rosszindulatú szereplők különböző technikákat használnak a SEO-mérgezés megvalósításához. Az egyik leggyakoribb módszer a **typosquatting**. Ezzel a módszerrel azok kerülnek célkeresztbe, akik a böngészőjükbe rosszul írják be az URL-címeket. Olyan domeineket hoznak létre, amelyek neve majdnem megegyezik az eredeti, másolni kívánt weboldallal, majd megvárják, hogy a felhasználók az ő oldalaira érkezzenek meg.

Examples of Typosquatting

Real Domain Targeted

[www.github.com](#)
[www.google.com](#)
[www.amazon.com](#)
[www.victoriass[□]ecret.com](#)
[www.homed[□]epot.com](#)

Typosquat Domain Example

[www.gIthub.com](#)
[www.gougle.com](#) Typos
[www.amozon.com](#) Missing an 'S'
[www.victoriasecret.com](#)
[www.homdepot.com](#) Missing an 'E'

3. ábra
Typosquatting példák

Ha azonban a typosquattingot SEO-mérgezéssel kombinálják, a rosszindulatú webhelyekre mutató linkek gyakran a keresési eredmények tetején jelennek meg, így a felhasználók nagyobb valószínűséggel kattintanak rájuk, mivel a címek nagyon közel állnak a valódiakhoz.



Blackhat SEO

A blackhat SEO a webhelytulajdonosok etikátlan taktikája, amelyeket a keresőmotorok rangsorának növelésére használnak.



Keyword Stuffing

Irreleváns kulcsszavak a weboldal szövegébe való rejtése. A meta címkékbe vagy a weboldal egyéb részeibe, hogy a keresőmotorok algoritmusait megtévesztve magasabb helyezést érjenek el a weboldalon.

Which Vacuum Cleaner is best for me?

Which is the **best vacuum cleaner** for you? With so many choices around, deciding for the **best model** may prove to be a difficult task.

The **most used vacuum** cleaners these days are the **upright and canister** types.

So if you are trying to choose the **best vacuum cleaner** among them, you need to keep a few things in mind.

4. ábra
Typosquatting példák

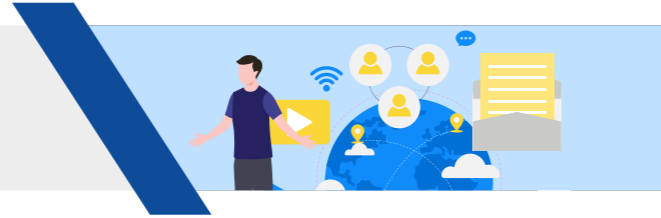
Cloaking

A keresőmotorok robotjainak más anyagot mutatnak be, mint ami a felhasználónak a linkre kattintáskor megjelenik. Ennek a módszernek a lényege, hogy amikor érzékeli, hogy egy keresőbot a „látogató”, számára legitim információkat mutat. Amikor egy emberi felhasználó kattint az oldalra, számára teljesen más, káros tartalmat jelenít meg.

A keresési rangsor manipulálása

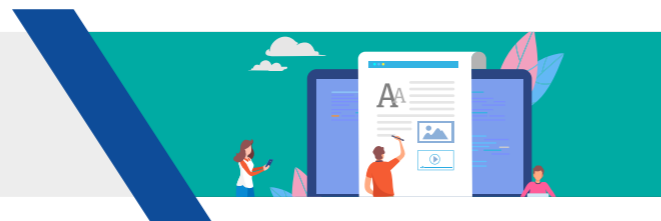
Egy weboldal átkattintási arányának mesterséges növelése érdekében a keresőmotorokban elfoglalt helyezésének növelésével. Ez a módszer botokat vagy embereket használ arra, hogy kulcsszavakra keressenek, és hamis kattintásokat generáljanak egy adott weboldal számára.

Private Link Networks

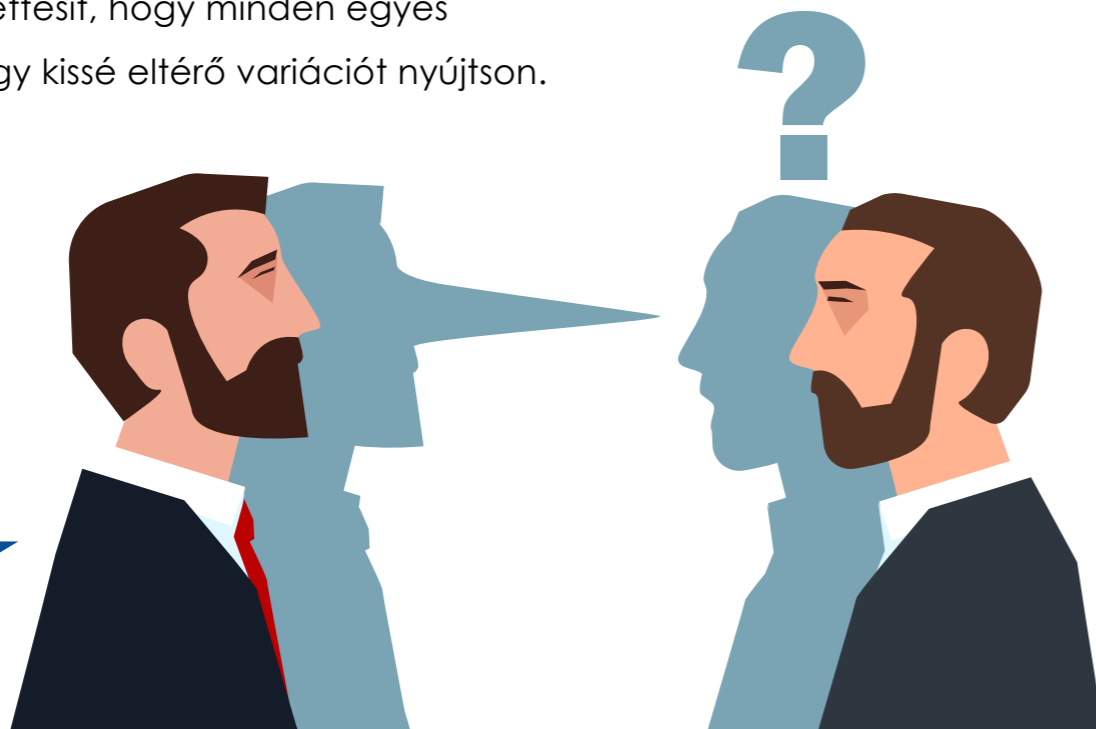


Nem kapcsolódó webhelyek csoportjának létrehozása és összekapcsolása egymással, ami egy fő webhelyre mutató backlinkek hálózatát eredményezi. Ez is egy módszer a keresőmotorok eredményeinek mesterséges növelésére, mivel a legitim linképítési gyakorlatok utánzására törekszik.

Article Spinning



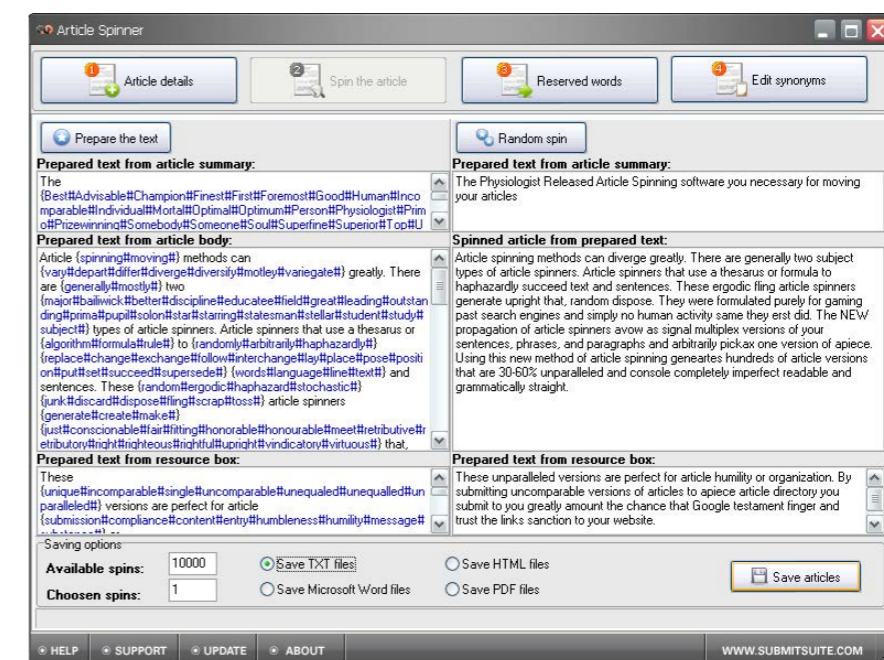
Egy olyan, a keresőmotor optimalizálásban (SEO) és más alkalmazásokban használt írásmód, amely a már meglévő tartalomból csalárd módon új tartalomnak tűnő tartalmat hoz létre. Az article spinning úgy működik, hogy bizonyos szavakat, kifejezéseket, mondatokat vagy akár egész bekezdéseket tetszőleges számú alternatív változattal helyettesít, hogy minden egyes változtatással egy kissé eltérő variációt nyújtson.



Ez a folyamat manuálisan megírható vagy akár teljesen automatizálható is. Az automatizált módszerekkel előállított korai tartalmak gyakran nehezen vagy egyáltalán nem olvasható cikkeket eredményeztek. Az article spinning technikák tökéletesedésével azonban egyre kifinomultabbá váltak, és ma már olyan olvasható cikkeket eredményezhetnek, amelyek felületes áttekintés után eredetinek tűnhetnek.

A webhelyek szerzői az article spinninget azért alkalmazzák, hogy csökkentsék a meglehetősen felesleges vagy minimális, illetve értelmetlen vagy nem informatív tartalmú oldalak hasonlósági arányát, és hogy elkerüljék a duplikált tartalom használatáért járó büntetéseket a keresőmotorok találati oldalain.

A végeredmény több hasonló, de különbözőnek álcázott dokumentum.



5. ábra
Article Spinning program működése

Sneaky Redirects



Az átirányítás az a cselekmény, amikor a látogatót az eredetileg kért URL címtől eltérő URL címre küldik. Bár nem minden átirányítás rosszindulatú, egy fenyegető szereplő vásárolhat olyan domainnevet, amely egy domain elírt változata. A fenyegető szereplő ezután a hibásan írt domainre véletlenül navigáló felhasználókat egy észrevétlen 302-es átirányítással a csalárd webhelyre irányítja át.

- ✓ https://mybank.com/user/login?redirect_url=/account/balance
- ✗ https://mybank.com/user/login?redirect_url=https://nybank.com/account/balance

6. ábra
Sneaky Redirect példa

SEO spam vagy spamdexing



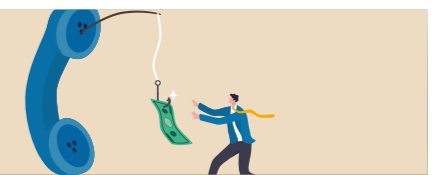
A keresőmotorok rangsorolásának manipulálására irányuló technika, hogy a forgalmat a fenyegető szereplő domainjére csalogassák. Ennek érdekében a hackerek hozzáférést szereznek egy normál, legitim weboldalhoz, majd kulcsszavakat és linkeket juttatnak be a webhelyeikhez amelyet affiliate marketing, a keresési forgalom pénzzé tétele vagy más rosszindulatú tevékenység céljából hoztak létre.

A spamdexing magában foglalhatja a kulcsszó vagy meta tag stuffingot, a beillesztett linkeket, és a sneaky redirects-et is.

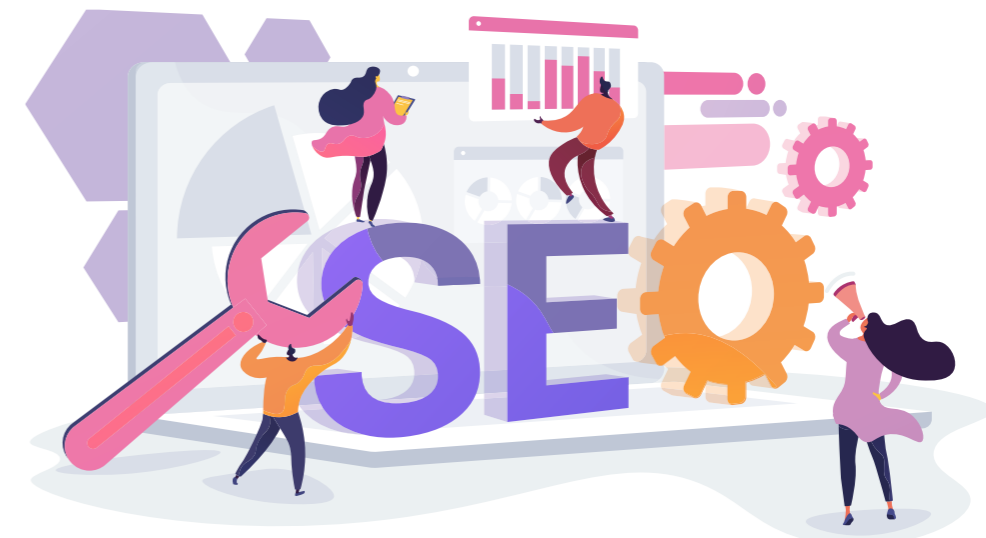
Ahelyett, hogy a legtöbb legális weboldalhoz hasonlóan rangsorolnának, a fenyegető szereplők egy normális weboldal hitelességét használják ki. Ez egy rövid életű stratégia, mivel a keresőbotok előbb-utóbb rosszindulatú szoftverek miatt észlelik az oldalt, amely káros tevékenységet folytat és feketelistára helyezik azt.

A hackerek azonban gyakran több ezer webhelyen hajtják végre ezt a támadást, hogy maximalizálják a hozamukat.

SEO scam



Olyan csalárd technika, amelyet a digitális marketingcégek és a szerződött SEO-tanácsadók hoztak létre, hogy a vállalkozókat és az üzlettulajdonosokat átverjék. Olyan keresőoptimalizálási szolgáltatásokért fizetnek, amelyet végre sem hajtanak, vagy nem hoznak tényleges hasznot.



Egyéb technikák

Injection



A támadók a weboldalak bővítményeinek sebezhetőségeit veszik célba. Kihasználják a nemrég nyilvánosságra hozott sebezhetőségeket, és külső szkripteket injektálnak. A hackerek ezzel a technikával csalárd weboldalakat adhatnak az eredeti oldalhoz, káros letöltési hivatkozásokat adhatnak hozzá, vagy más oldalakra irányíthatnak át.

Siteurl hackek



Egyes sebezhetőségek a weboldalak tetszőleges opcióinak módosítását teszik lehetővé, ami úgynevezett siteurl fertőzéseket eredményezhetnek. Ebben az esetben a weboldal maga kezdi el betölteni az összes erőforrást, például a stílusokat és a szkripteket a rosszindulatú URL-ekről.



HTML injekciók



Bizonyos esetekben a sebezhetőség lehetővé teszi egyéni HTML-kód befecskendezését egy bővítmény vagy téma opciójába. A webhelyhez való további hozzáférés érdekében a befecskendezett szkriptek gyakran tartalmaznak olyan kódot, amely megpróbálja rávenni a fertőzött webhelyek bejelentkezett adminisztrátorait, hogy hajtsanak végre parancsokat az admin felületen (pl. új admin felhasználók létrehozása).



Adatbázis injektálások



Ha a hackerek ellopják weboldal adatbázis hitelesítő adatait a config fájlból, lehetővé teszi a támadók számára, hogy ezután rosszindulatú programokat injektáljanak az összes közzétett bejegyzésbe és oldalba. Emellett a támadók további admin felhasználókat is létrehozhatnak.



Tetszőleges fájlbefecskendezések



Néhány súlyos sebezhetőség lehetővé teszi a támadók számára, hogy tetszőleges fájlokat töltsenek fel a veszélyeztetett szerverekre, vagy tetszőleges kódot hajtsanak végre.

A Sucuri egyik elemzője nemrégiben egy új Mal.Metrica átirányítási csalást azonosított. A fertőzött webhelyen egy (hamis) emberi hitelesítési kérés jelent meg:



Please verify you are a human

PRESS

Access to this page has been denied because we believe you are using automation tools to browse the website.

This may happen as a result of the following:

- Javascript is disabled or blocked by an extension (ad blockers for example)
- Your browser does not support cookies

Please make sure that Javascript and cookies are enabled on your browser and that you are not blocking them from loading.

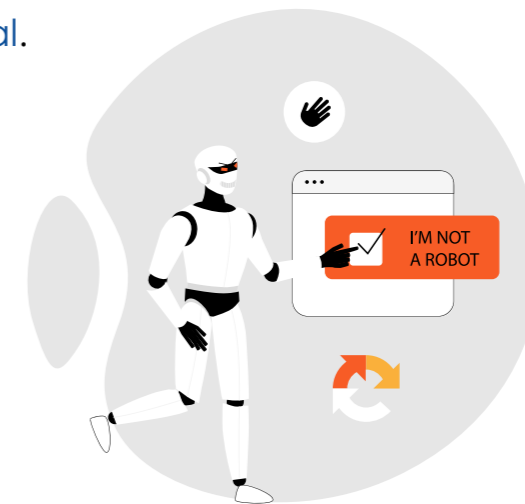
Reference ID: #4675a200-f2b8-11ea-bd91-6912832c013d

7. ábra
[Egy hamis hitelesítési captcha](#)

Ezek a kérések manapság **elég gyakoriak** a weboldalakon, és a legtöbb felhasználó valószínűleg nem is gondolkodna, hogy rákattintson-e. Mindenki tudja, hogy mennyi időt töltünk azzal, hogy tűzcsapokra, buszokra és közlekedési lámpákra kattintva a Google CAPTCHA ellenőrző kérésekben bizonyítsuk, hogy emberek vagyunk.

Bár ez a felszólítás **rutinszerű ellenőrzésnek tűnik, valójában teljesen hamis** - és ehelyett megpróbálja becsapni a felhasználót, hogy kattintson a gombra, ezáltal rosszindulatú és csaló weboldalakra történő átirányítást kezdeményezve.

Ahelyett, hogy JavaScriptet injektálna a webhely kódjába (ami nagyon gyakori a rosszindulatú programok befecskendezésénél), a fertőzés egyszerűen **egy képfelületet hoz létre egy rosszindulatú domainre mutató hivatkozással**.



A **mesterséges intelligencia**, a fejlődésének és térhódításának köszönhetően, **már ezekhez a csalásokhoz is felhasználható**. Az article spinninghez pillanatok alatt elkészít több variációt, így megkönnyítve a csalók dolgát, de képes akár komplett weboldalakat, nem létező személyek képeit is legenerálni rövid idő alatt.

Az [Engadget](#) számolt be egy csalási kísérletről, amely során egy kiadó egy csalárd "szerzői jogi jogsértési értesítés" e-mailt kapott egy ügyvédi irodától. A feladó egy "látható és kattintható link" hozzáadását követelte a weboldalához az állítólagos lopott fotó alatt, azonban a kiadó tudta, hogy egy jogdíjmentes szolgáltatótól származó fotót használt, ezért utánanézett a követelésnek, és kiderült, hogy az ügyvédi iroda ügyvédjeinek képeit mesterséges intelligencia generálta.

A képkérés a „generated.photos” nevű weboldalhoz vezetett, amely mesterséges intelligenciát használ modellfotók létrehozására.

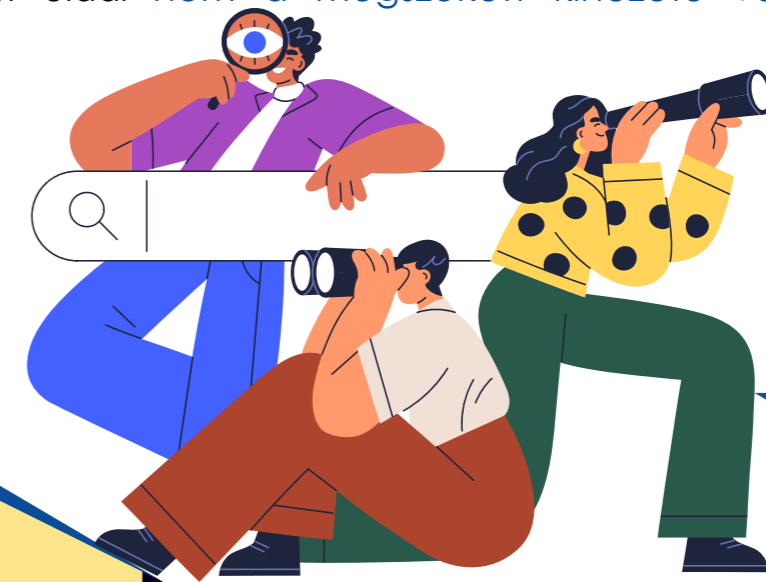


Összegzés

Egy keresőmotor lényege, hogy a felhasználók által beírt kulcsszavak alapján találatokat jelenítenek meg, amelyek a lehető leginkább releváns oldalakra vezetik az internetezőket a keresési szándékuknak megfelelően. Minden keresőmotornak megvan a saját algoritmus, amely alapján az indexelt és feltérképezett weboldalakat rangsorolja.

A nyelvi modellek előre törésével egyre nagyobb hangsúlyt kapnak az ilyen jellegű keresések, amely a hagyományos keresők rovására fog menni idővel. Míg pl. egy Google kereséssel egy releváns oldalt kaphatunk meg, addig a nyelvi modell által generált eredmény teljes körű választ nyújthat a kérdésünkre. Az AI alapú keresés sokkal több gépi tanulást épít be, mint a versenytársak, de ennek is megvannak a maga kihívásai, különösen a természetes nyelv feldolgozása terén.

Érdemes óvatosságnak lennünk, hogy mire kattintunk, és óvakodjunk minden gyanús dologtól, ami a böngészőnkben felbukkan - még akkor is, ha az egy olyan weboldaltól érkezik, amelyben egyébként megbízunk. Ellenőrizzük a látogatott weboldal URL címét, fogjuk gyanút, ha a meglátogatott oldal nem a megszokott kinézetű vagy információ tartalmú.





NEMZETI
KIBERVÉDELMI INTÉZET



nki.gov.hu



titkarsag@nki.gov.hu



+36 (1) 325 7672



Nemzeti Kibervédelmi Intézet



@ [nki.gov.hu](https://www.instagram.com/nki.gov.hu)



Kibertámadás!
podcast