

**Debreceni Egyetem
Informatikai Kar**

Egy Intelligens Web-kereső

Témavezető:
Mecsei Zoltán
számítástechnikai munkatárs

Készítette:
Csekő Zsolt
programozó matematikus szak

DEBRECEN
2008.

TARTALOMJEGYZÉK

1	BEVEZETÉS.....	2
2	A WEB KERESŐK TÖRTÉNETI ÁTTEKINTÉSE	4
	<i>Papír alapú keresés, Manuális keresés.....</i>	<i>5</i>
	<i>Az online információkeresés kezdetei.....</i>	<i>6</i>
	<i>A második generációs keresőprogramok újításai a keresési eredmények megjavítására:.....</i>	<i>8</i>
	<i>A kereső rendszerekről manapság</i>	<i>8</i>
3	A KERESŐRENDSZER TARTALMA	9
	<i>Keresési lehetőségek.....</i>	<i>9</i>
	<i>A keresés eredménye.....</i>	<i>10</i>
	<i>Teljesítményértékelés</i>	<i>10</i>
	<i>Keresőoptimalizálás, fokozatai, célja</i>	<i>10</i>
4	A KERESŐRENDSZEREK JÖVŐJE.....	13
	<i>Indexelés és visszakeresés.....</i>	<i>13</i>
	<i>Felhasználói beállítások</i>	<i>13</i>
	<i>Összesítés</i>	<i>13</i>
	<i>A keresés kiterjesztése, hatókör</i>	<i>14</i>
	<i>Metaadatok</i>	<i>14</i>
	<i>Következtetések</i>	<i>14</i>
5	A WEB KERESŐKRŐL ÁLTALÁBAN (KERESŐROBOTOK)	15
6	EGY INTELLIGENS WEB-KERESŐ.	21
7	KONKLÚZIÓ	33
8	KÖSZÖNETNYILVÁNÍTÁS	34
9	IRODALMI JEGYZÉK	35

1 Bevezetés

A gyökerek a hatvanas évekig nyúlnak vissza, hadifejlesztésként indult, így a katonai fejlesztések civil szférába való átszivárgásával kezdődött. Abban az időben, a hidegháborús időszak idején merült föl ugyanis az USA-ban egy kevésbé sebezhető számítógép-hálózat szükségessége, amelynek egy esetleges atomtámadás után megmaradó központjai és alközpontjai továbbra is fenn tudják tartani egymással a kapcsolatot, vagyis Amerika szervezett és irányítható maradjon. Abból indultak ki, hogy egy országos információs és irányító hálózat egyetlen központja elsődleges célpontja lenne a támadásnak, tehát azonnal megsemmisülne. A megoldás, a decentralizáció. Olyan rendszert kell tehát létrehozni, amelynek nincs egyetlen kitüntetett központja, hanem eleve kis alegységek formájában működik. Fontos követelmény, hogy a keletkező struktúra szabadon konfigurálható legyen abban az értelemben, hogy új csomópontok felvétele illetve, eltávolítása egyszerűen elvégezhető műveletek legyenek, de akár néhány csomópont megsemmisülése se legyen katasztrofális a rendszer egésze szempontjából.

A 60-as évek második felében a RAND Corp., a Massachusetts Institute of Technology (MIT) és az University of California at Los Angeles (UCLA) kísérletezett a csomagokra bontott információ átviteli módszereinek a kifejlesztésével. Az első próbahálózatot a National Physical Laboratory brit intézet hozta létre 1968-ban. Majd a Pentagon hatáskörébe tartozó Advanced Research Project Agency (ARPA) is bekapcsolódott a kísérletekbe. Az ARPA kutatói, olyan rendszert képzeltek el, amelynek csomópontjait nagyteljesítményű szuperszámítógépek alkotják. A tervezés és kivitelezés során gondoltak arra, hogy ez a hálózat békeidőben is kitűnő lehetőséget teremthet egymástól távol eső erőforrások - adatbázisok, számítási kapacitás, stb. - elérésére.

A világháló (angol eredetiben World Wide Web, WWW vagy röviden Web) az interneten működő, egymással úgynevezett hiperlinkekkel összekötött dokumentumok rendszere. A rendszert webböngésző program segítségével lehet elérni. Ez a program képes megjeleníteni az egyes dokumentumokat, „weblapokat”. A felhasználó a lapokon található hiperlinkek segítségével további lapokat kérhet le, amelyeken újabb hiperlinkek lehetnek.

A rendszer „háló”-jellegét is ez adja; a dokumentumok a háló csomópontjai, míg a hiperlinkek a háló szálai, amelyeken keresztül egy vagy több lépésben tetszőleges

csomópontokhoz eljuthatunk. Az internet robbanásszerű elterjedése maga után vonta a visszakereshető információ mennyiségének exponenciális növekedését.

A nem várt mértékű növekedés, valamint a rendszertelenség hamar szükségessé tette a digitális keresés gyors fejlesztését. Számos, internetes keresőmotor hatalmas mennyiségű metaadatot tárolt.

Az internetes keresőrendszerek hatalmas fejlődésen mentek keresztül az utóbbi két évtizedben, és a fejlesztők szerint még jelentős előrelépésekre van lehetőség. A kutatások több különböző irányban haladnak, az eltérő megközelítések által mutatott utak jelenleg még nem határozzák meg, hogy mely elméletek válnak be hosszútávon a gyakorlatban.

Alapvetően két nagy típusát különböztethetjük meg a keresőgépeknek: az egyik esetben emberek által szerkesztett listákon, könyvtárakon alapuló keresésről beszélünk, míg a másik esetben úgynevezett webrobotok (automatikusan indexelő keresőgépek) irányítják a keresést.

Hogyan működnek a keresők? Az első esetben könnyebb a működés menetét áttekinteni. Az adatbázisokat, listákat emberek szerkesztik, a webhely tulajdonosának kell kezdeményeznie az oldalára mutató link felvételét a listára, azaz regisztrálnia kell magát. Általában meghatározhatja, mely kategóriában akar jelen lenni. A listát annak felelősei tartják karban.

Az automatikusan indexelő keresőgépek esetében már egy kissé összetettebb folyamatról beszélhetünk. Az úgynevezett "pók" kezdeményezi az ismerkedést, azaz meglátogatja a weblapokat, olvassa őket, és követi a más oldalakra mutató linkek útját. Azonban a pókok tevékenysége nem csak a kapcsolatfelvételt jelenti, mivel bizonyos időközönként (általában 1-2 hónap) visszatérnek a weboldalakra és megnézik történt-e változás.

A következő fázisban a pókok összegyűjtik, más szóval indexelik találataikat, kereséseik eredményét. Összegyűjtik, besorolják egy hatalmas katalógusba (index), azaz "szövik hálójukat". Ha e katalógus a valós világban megjelenne, olyan gigantikus méretű lenne, hogy nem is tudnánk hova elhelyezni, mivel e katalógus a pókok által feltárt összes webhely oldalait tartalmazza.

Ha egy oldal megváltozik, az automatikusan bekerül e katalógusba, így módosul folyamatosan és dinamikusan a tartalma. Azonban időbe kerül, míg a pók észleli a változást és hozzáadja a katalógushoz.

2 A web keresők történeti áttekintése

A 80-as évek végétől vált lehetővé az interneten való keresés széles körben való elterjedése. A telnet-től a webig vezető úton számos új módszer jelent meg az elérhető információk áttekintésére. A gyors fejlődést már nem lehetett nyomon követni csupán szóban, elektronikus levelezéssel vagy sokszorosított listák útján.

A kezdeti eszközök a laikus felhasználó számára elég nehézkesen használhatók voltak. Ilyen az FTP (File Transfer Protocol), az adatátvitel általános szabványa, s egyben program, ami biztosítja, hogy a hálózat számítógépei között egységes formában jöjjön létre az állománycsere. A WAIS (Wide Area Information Servers), az Archie hozták létre a másodlagos adatállományokat (indexek, relevanciaadatok, leírások), amelyek az internetre kerülő dokumentumokra utalnak. Ezek a szerveroldali indexelő rendszerek. Ezek a háttérben működnek, ezért használatuk, mint már említettem, meglehetősen bonyolult.

Az áttörést két rendszer megszületése hozta: 1991-ben Gophereket készítettek. Ez egy szöveges, menüszerkezetű információs hálózat, ami hasonlít a hierarchikus tartalomjegyzékekhez. "A Gopherben a felhasználó szöveges (és csak szöveges) állományokat nézhetett meg és tölthetett le. Legismertebb keresőeszköze a Veronica integrált menülelkérdező és indexelő rendszer (Very Easy Rodent-Oriented Net-wide Index to Computerized Archives)."

Ugyancsak 1991-ben Tim Berners-Lee kísérleteiből kiindulva született meg a World Wide Web hipermédia információs hálózat üzemszerű formája.

"Ebben a rendszerben a hypertext jóvoltából az információforrásokra hypertext csatolók (hyperlinks) formájában teljesen szabadon lehet egymással összekapcsolni. A kapcsolódó és megjeleníthető információforrások nemcsak szövegek, hanem képek, hangok és mozgóképek is lehetnek. A rendszeren belül nem érvényesül semmiféle hierarchikus rendező elv, minden forrás egyszerre több forrással is összefügghet és fordítva (azaz a kapcsolódások szerkezete M:N). A weben belül hamarosan kialakultak különféle keresőszolgáltatások, mint például a Galaxy, az Infoseek, a Yahoo!, a Lycos vagy az AltaVista.

A web formájában végre megvalósult az általánosan hozzáférhető és asszociatív gondolkodással összhangban álló felhasználói felület, amelyet 1945-ben Vannevar Bush

megálmodott. Benne minden addigi egységesítő (FTP), feldolgozó és keresőeszközt (WAIS stb.) integráltak."

A cél tehát nem más, mint megkeresni valamit az interneten. Ez nem egy egyszerű feladat, több milliárdnyi oldal közül kell kiszűrni, amit meg szeretnénk valahogy találni. Erre több módszer is van, az egyik emberek által összeállított linkgyűjteményekre épít. Ezeknél általában sokkal kevesebb lap található meg, főként népszerűbb témákban, viszont sokkal jobb minőségű, egyszerűbben áttekinthető. A feladat nehézségét érzékeltetendő, gondoljunk bele, mekkora munka lehet a milliárdnyi weblap kategorizálása, főként, hogy naponta születnek újak, szűnnek meg régiek! Ebből adódik, hogy kevesebb -viszont az emberi intelligencia miatt nagyon jól az adott kategóriába illő- weblapot találhatunk ilyen helyeken. A másik lehetőség a keresőrobotok által folyamatosan pásztázott oldalak alapján készített hatalmas adatbázisok. A kis robotprogramok folyamatosan kéri le a weblapokat, feldolgozzák, megpróbálják megállapítani, hogy miről is szól, majd ezt eltárolják magukban. Ennek az adatbázisnak a felhasználói felületét hívjuk keresőnek. Ezeknél egy fokkal nehezebb hozzájutni az adatokhoz, hiszen nem emberi feldolgozáson estek át, előfordul, hogy hibás, vagy legalábbis nem olyan találatokat kapunk, amire számítottunk. Persze ehhez jól is kell kérdezni, hogy minél kevesebb hibás és minél pontosabb jó találatunk legyen. - azt fogjuk áttekinteni a továbbiakban, hogy ez hogyan is zajlik.

Papír alapú keresés, Manuális keresés

A hagyományos rendszerű könyvtári katalógus-rendszer indexelt kartotékok segítségével tette, és teszi még most is átláthatóvá és kereshetővé a kívánt művet. A kartotékokra felvitt különböző kulcsszavak segítségével bonyolult, összetett kapcsolatok hozhatók létre, melyek elősegítik a kifinomultabb keresések megvalósítását. A technika legnagyobb hátránya a korszerűbb, elektronikusan segített keresésekhez képest az, hogy a felhasználónak pontosan tudnia kell, mire is van szüksége, különben igen körülményes a megfelelő mű beazonosítása. A könyvtári katalógusok elektronikussá válásával előfordulásuk eltűnőben van, mindazonáltal számos szakcikk továbbra is csak ezen a módon érhető el. A dokumentumok teljes digitalizálására lehet, hogy soha sem nem kerül sor, de az őket leíró, katalogizáló rendszerek már gyakorlatilag mindenhol elektronikusak vagy hamarosan azok

lesznek. Elmondható, hogy a hagyományos keresés kiszorulóban van, csökkenő tendenciát mutat a csak így fellelhető művek száma – még, ha tanulmányozásuk esetleg továbbra is kizárólag nem elektronikus formában történhet.

Az online információkeresés kezdetei

A tartalom szerinti információkeresés fejlődése már a II. világháború után elkezdődött, és a hypertext megjelenésén keresztül vezetett el az interneten megvalósult világméretű információkereséshez. Vannevar Bush, aki a háború alatt az amerikai tudósok "hadseregét" irányította, 1945-ben fogalmazta meg először, hogy az információkeresésnek asszociatív kapcsolatokon kell alapulnia (ekkor még nem használták az "information retrieval" kifejezést). Ő használta először az összekapcsolt szövegblokkok fogalmát, ő vezette be a 'link' (ebben az esetben a *releváns* szöveghelyre utaló egyszerű kapcsolatjelölő: csatoló, kapocs, utaló, mutató, hivatkozás, ugrópont) és a nyomvonal, valamint a háló kifejezéseket a textualitás új elképzelésének a leírására. Koncepciója a gépesített, határtalan kapacitású, mindenféle dokumentumokat tartalmazó iratgyűjtemény és könyvtár, amely a felhasználó számára gyors, asszociatív keresést tesz lehetővé. A fél évszázada megálmodott elektronikus, hálózati könyvtár feltételei napjainkra értek meg.

Vannevar Bush munkássága nagy hatással volt többek között Theodor Holm Nelsonra. Nelson írta le először a hypertext nevet 1965-ben, és meghatározta fogalmát felhasználói szempontból: "Írott vagy képi anyagok olyan komplex összeköttetése, amit papíron nem lehet kényelmesen megalkotni. Összefoglalókat és térképeket tartalmazhat a benne szereplő anyagokról és ezek egymáshoz való viszonyáról; tartalmazhatja az anyaggal foglalkozó tudósok megjegyzéseit és lábjegyzeteit is.

A számítógépes adatbázisok első pillanatától fogva szükség volt az adatok gyors, hatékony kikeresésére. A szűkös tárhelykapacitás, a drága gépidő jelentősen korlátozta az első keresőrendszerek hatékonyságát. Mindazonáltal ne feledjük el, hogy erre a korszakra tehető a rendszerek elfogadottá válása, széleskörű elterjedésük a különböző – üzleti, katonai – területeken.

Az első generációs online adatbázisok jelentősen továbbléptek ma már primitívnek számító elődeikhez képest. Jelentőségük abban nyilvánult meg, hogy lehetőséget adtak arra, hogy egy rögzített rekord minden egyes mezőjében bármely kulcsszó szerint lehessen keresni.

Mindezen túl bevezették a logikai parancsszavak – AND, OR, NOT – használatát. A fejlődés természetesen nem állt meg ezen a ponton. Ezzel egy időben jelentek meg a keresőkben azok a lehetőségek, amelyeknél a felhasználó már az egyes rekordmezők típusa szerinti kereséseket is kiadhatott (médiatípus, dátum, nyelv, cikk vagy tanulmány, stb.).

A keresési lehetőségek ilyen mértékű elburjánzása – nyilvánvaló előnyei ellenére – a rendszert használók számára egyre nehezebbé tette magának a keresőnek a használatát. Mivel az egyes keresőmotorokhoz definiált kulcsszavak semmilyen szabványt nem követtek, minden egyes keresőrendszer használatához új 'tájnnyelvet' kellett megtanulni. (Ne felejtjük el, hogy a napjainkban használt grafikus felhasználói felületek ekkor még nem léteztek, a kereséseket parancssorban, kézzel kellett megadni – minden beállítással egyetemben.) Az ekkor használt protokollok és archaikus nyelvek használatát az sem könnyítette meg igazán, hogy az online segítség használata is roppant körülményes volt

Érdekes velejárója volt a fejlődésnek, hogy bár számos adatbázis használata ingyenes volt, a keresés maga már nem. A keresés költségét a keresőmotort üzemeltető számítógép gépidejétől tették függővé. Ennek számos bosszantó következménye lehetett. Például egy melléüötött karakterrel kiadott hosszantartó keresés mindamelllett, hogy sikertelen volt, nem kevés pénzbe is került. Hamar felmerült az igény a felhasználóbarátabb környezetek megteremtésére.

A Windows – grafikus – környezetek elterjedésével hamarosan megjelentek a fel- és legördülő menüs, listás és mezős felületek. Sajnos ezek egymással nem voltak kompatibilisek, így a használatukból adódó előnyök csak korlátozott mértékben jelentkeztek. (Néhány korabeli felület kifejezetten ronda volt, míg másoknak a sebessége volt az elfogadható kritérium alatt.)

A CD-ROM megjelenésével és tömeges elterjedésével azonban a keresőrendszerek átléptek egy korlátot. Az eddig viszonylag szűk, tudományos körökben használt számítógéppel felgyorsított keresés egy csapásra elérhető lett, és egy viszonylag olcsó 'offline' médium segítségével mindenki számára megfizethető módon kezelhető körbe került. Az ilyen típusú adatbázisok legjellemzőbb és kézzelfogható példája a jogi adattár.

Ennek CD-re kerülése számos jogi kérdés eldöntését több nagyságrenddel felgyorsította, a változó jogszabályok követése pedig egy egyszerű frissítéssel megoldható. A rendszer mindenki számára kényelmes, praktikus és nem utolsó sorban olcsó. Sikerességét mi sem jellemzi jobban, minthogy az online média korában is életképesnek bizonyult, még jó pár év telik el mire teljesen kiszorul a piacról.

A második generációs keresőprogramok újításai a keresési eredmények megjavítására:

A keresés irányultságának fogalmi felismerése. Ide tartozó területek az alkalmazott jelentéstan, a természetes nyelvi feldolgozás. E technikákat alkalmazó két legjelentősebb kereső az Ask Jeeves és a Northern Light.

Az Ask Jeeves adattárában számos, különböző tárgyterületek szakértői által előre megfogalmazott kérdés van, amelyekhez hozzárendeli a legrelevánsabb válaszoldalakat. A kereső elemzi a felhasználó által feltett kérdést, majd hozzárendeli a hozzá legjobban hasonlító előredefiniált kérdéshez, s szolgáltatja a válaszokat.

Northern Light a keresés eredményét a természetes nyelvi feldolgozó alkalmazásával csoportosítja site és/vagy tartalom alapján, s a felhasználó e csoportok közül választhat.

Hivatkozások analízise: A Google elsősorban a dokumentumra mutató külső hivatkozásokkal számol, s a népszerűbb oldalról történő külső hivatkozásokat súlyozottan veszi figyelembe. Algoritmusai annyira hatékonyak, hogy általánosan elfogadottan a Google-t tartják a legjobb keresési eredményeket nyújtó szolgáltatásnak.

Az oldal népszerűségének elismerése: Számos kereső az eredménylista sorrendjének meghatározásánál figyelembe veszi az eredményoldalak népszerűségét. Pl. Google, DirectHit.

A DirectHit egy oldalt azon elv alapján rangsorol, hogy egy egyszerű keresés eredménylistájából hányan választják az adott oldalt.

A második generációs keresőprogramokra jellemző, hogy egy oldal helyezését a rangsorban előnyösen befolyásolja az oldal népszerűsége. Ha ehhez hozzávesszük, hogy a felhasználók ritkán böngésznek a keresési eredménylista második oldalán túl, akkor az a várható trend, hogy az eddig is népszerű oldalak még népszerűbbek lesznek, az új oldalaknak pedig egyre nehezebb lesz jó helyezést elérni.

A kereső rendszerekről manapság

Két alaptípus van: tárgyszótárak (Yahoo) és lekérdező robotok (AltaVista, HotBot, Excite). Mindkettő weblapokra mutató adatbázist tart karban. A tárgyszótárak valamilyen – gyakran meglehetősen bonyolult hierarchiájú – osztályozási rendszer szerint csoportosítják a

forrásokat. Az automaták a kereső személy által beírt szöveges kifejezésre épülő algoritmust futtatnak. Mindkét típus alkalmaz a másik logikájából átvett kiegészítő módszereket.

A hálón keresgélő informatikusok próbálnak megbirkózni a túlburjánzott általános és speciális keresőrendszerekkel: osztályozott keresőmotor-jegyzékeket és a keresőkérdést a többi konkrét keresőhöz továbbküldő, a bejövő eredményeket összefésülő metagépeket bocsátanak közre.

3 A keresőrendszer tartalma

Keresési lehetőségek

A weboldal kétféleképpen kerülhet fel a keresőrendszer indexébe: vagy a gazdája jelenti be, pl. az erre kibocsátott weblapon, vagy a keresőrendszernek a linkeket végigkövető, illetve az új és változott lapokat felkutató ügynökszoftverje, robotja deríti és jegyzi fel URL-jét. Az ügynökszoftverek kifinomult stratégiával súlyoznak a bejárható útvonalak között. Az AltaVista pl. napi 6 millió lapot gyűjt be.

A tárgyszótáras rendszerek a felderítést igyekeznek a lényeges forrásokra korlátozni, és a felvétel előtt szűrik a begyűjtött információt.

Az indexelés szolgálatról szolgálatra, más elemekre terjed ki az URL-től a teljes szövegig. Az ismertetést olykor kézi úton kulcsszavakkal, osztályozással stb. egészítik ki, ami a minőséget javítja, de a forgást lassítja.

A tipikus nyitó képernyőn van egy input box; a beírt szavakon felül az egyén bejelölhet bizonyos szűkítéseket. Sok keresőgép kínál ínycsekknek való keresőparancsokat, de az eseti felhasználó jobb, ha óvakodik ezektől, pl.:

- Boole-algebrai kifejezések (esetenként zárójelezéssel);
- azoknak a szavaknak a megadása, amelyeknek benne kell lenniük, vagy amelyeknek nem szabad benne lenniük a keresett dokumentumban;
- csonkolás;
- teljes kifejezésre keresés;
- helyzeti operátorok;
- korlátozás adatmezőre;
- kis- és nagybetűk megkülönböztetése;
- szűkítés: dátum, nyelv, terület, fájltypus szerint.

A keresés eredménye

A kapott találatok sokféle formális és tartalmi ismerv (pl. a keresőszavak előfordulásának gyakorisága, vagy az előfordulás helye, vagy a keresőkérdésbeli súlyozás) szerint rendezve jeleníthetők meg. A keresőgépek egyszerre kb. 10 találatot mutatnak meg, és kijelzik a teljes találati halmaz méretét, amiben azután navigálni lehet.

A megjelenés formátumát is szabályozni engedik, sőt megkereshetik a találatokban leggyakoribb szavakat, hogy a következő iteratív lépésben ezzel finomítsuk a keresést.

Teljesítményértékelés

Bőségesen léteznek összehasonlító értékelések, de gyorsan elavulnak. A web gyakran aktualizált sajátosság-egybevető táblázatai már jobbak, ám nincsenek elég jól specifikált mérési módszerek. Professzionális mérések is történtek. A vizsgálatokban több rendszer szerepelt: AltaVista, Excite, HotBot, Infoseek, Lycos, Magellan, Open Text, Point. Ezeknek egyszerre tették fel ugyanazokat a kérdéseket, és vizsgálták a keresés sebességét, a találatok mennyiségét, relevanciáját. Az eredmények csupán tájékoztatóak, hiszen eleve nem lehetett nagy a kísérletek száma.

Végül is a vezető termékek között nem volt lényeges eltérés, és valahol mindnek volt speciális előnye is, ráadásul ezek gyorsan fejlődnek, hamar elavulttá téve minden összehasonlítást.

Keresőoptimalizálás, fokozatai, célja

A keresőoldalak felől érkező látogatók számának növelése a keresők rangsorolási szempontjait figyelembe vevő webfejlesztés segítségével történik. Mivel számtalan módon kötődik az online marketing egyéb ágaihoz, ezért sokan „*keresőmarketing*”-nek is hívják ezt a tevékenységet.

Az optimalizálás azon alapul, hogy a *szabadszavas keresők* csak közvetítők az interneten kereső felhasználók és a weboldalak üzemeltetői között: úgy is felfogható ez a viszony, hogy a keresők ingyenes hirdetési felületet biztosítanak a webmesterek számára, cserébe pedig a webmesterek által szolgáltatott információk választ adnak a felhasználók kérdéseire. A keresők elemi érdeke, hogy azokat az oldalakat hirdessék, amelyek a legjobb, legaktuálisabb, leghasználhatóbb választ adják a *kulcsszavakkal* megadott kérdésekre, mert

így nő a felhasználók elégedettsége, és ezzel saját látogatottságuk is. (ez pedig az interneten egyszerűen pénzre váltható, lásd: Pénzkeresés).

Kevés olyan felhasználó van, aki nem lép tovább a keresőoldalakról: a kereső rendszerek ezért az általuk generált látogatottság döntő hányadát egyenesen továbbadják a *találati rangsorban* megjelenő webmestereknek, akik aztán szintén profitálhatnak a látogatottságból. Azok szerzik meg a legtöbb ügyfelet vagy hirdetési bevételt, kiknek oldalai sokszor, jó helyen és megfelelő kivonattal jelennek meg; tekintve, hogy ez a három alapvető tényező befolyásolja az átkattintások arányát.

- Ha a webhely oldalai gyakran keresett kulcsszavakat tartalmaznak, akkor az adott webhely sokszor jelenik meg a keresők által nyújtott "hirdetési felületen".
- Minél jobban megfelel egy weblap a keresők *rangsorolási kritériumainak*, a lap kivonata annál előkelőbb, szembetűnőbb helyen fog megjelenni a listában.
- Mennél ígéretesebbnek tűnik a weboldal kivonata, annál biztosabb, hogy rákattint a felhasználó, ha már egyszer a szeme elé került.

Ahhoz, hogy weboldalunk lapjai sokszor, jó helyen, és érdeklődést felkeltő módon jelenjenek az, alábbi lépéseket kell megtennünk:

1. Jól megírt weboldalak létrehozása: A keresők szöveges információt olvasnak, és szöveges linkeket követnek. Alapvető fontosságú ezért, hogy mennyi és mennyire szabatosan megfogalmazott szöveg található az oldalon; továbbá, hogy minden szöveges információt el lehessen érni szöveges, Teljes értékű linkekkel. Ez a lépés pusztán oldalunk jobbá tételét jelenti, még nem is a weblap optimalizálás kategóriájába tartozik, mégis a későbbi keresőoptimalizálás elengedhetetlen feltétele.
2. Linknépszerűség növelése: Minél több helyről mutat Külső hivatkozás a webhelyre, annál több látogató érkezik, kivéve, ha Rossz szomszédságba keveredett a websájt. Csupán a linkek számának növelése azonban még nem optimalizálás, hiszen a linkek szaporítása nemcsak közvetlen látogatottság-növekedést generál: A bejövő hivatkozások a webhely Általános fontosságának növelésével közvetve is hozzájárulnak a keresők felől érkező forgalom erősödéséhez, a Linkre rakott kulcsszavak pedig az oldal Keresésfüggő fontosságának növekedéséhez járulnak

hozzá. Nem mindegy tehát, hogy honnan, a webhely mely részére és milyen szöveggel mutatnak linkek oldalunkra.

3. Webhely-optimalizálás: A weboldal célját tükröző szabatosan, jól megírt szöveg még nem feltétlenül tartalmazza az összes gyakran keresett kulcsszót, illetve nem feltétlenül azok a Kulcsszavak kerültek a leghangsúlyosabb helyekre, melyek a legtöbb látogatót hozzák. Ebben a lépésben tehát a kulcsszavak megfelelő kiválasztásáról és elhelyezéséről van szó. A Kulcsszavak kiválasztása során ki kell találni, hogy a webhely iránt potenciálisan érdeklődő [Céltudatosan kereső felhasználók] milyen kifejezésekre keresnek rá a leggyakrabban, és el kell dönteni, hogy ezek közül melyekre érdemes optimalizálni. Mivel a változtatások még mindig csak az optimalizált oldalt érintik, ezt a folyamatot szokták on-site optimalizáció néven is hívni.
4. Listamegjelenés-optimalizálás: Ha már előkelő helyen jelenik meg az optimalizált weboldal a megfelelő kifejezésekre rákeresve, az már fél siker. Ezt azzal fokozhatjuk még, ha úgy írjuk meg a weblapokat, hogy a találati rangsorokban megjelenő kivonatok figyelemfelkeltőek legyenek: felkeltsék a reménybeli látogatók érdeklődését (és egyszersmind ne elégítsék ki, ezzel bátorítva/kényszerítve őket az oldalra meglátogatására.)
5. Nagyipari kereső-optimalizálás: Ha egy vállalkozás nyereségessége és a weboldal látogatottsága között egyenes összefüggés van, akkor már nagyban kell játszani, ami sokszor magába foglalja a manipulatív, a Kereső programok ajánlásaival ellentétes technikák használatát.

Elsődleges cél tehát az, hogy a felhasználók kereséseinek minél nagyobb hányada végződjék az optimalizált oldalra mutató linke kattintással. Másodlagos cél, hogy minél több olyan látogatónk legyen, aki az optimalizált oldal célközönségébe tartozik. A kereső-optimalizálás tehát elsősorban arról szól, hogy minél több Céltudatosan kereső felhasználót vonzzunk oldalunkra. Rendszeres látogatókat vagy Szörfölő felhasználókat csak közvetve vonzhatunk a keresőoptimalizálás eszközeivel, ezért ha lehetséges, együtt kell alkalmazni a hagyományos reklámok, internetes marketing és a Keresőmarketing módszereivel. Mivel ez utóbbiak leginkább sok pénzbe kerülnek, sok webhely nem engedhet meg magának mást, csak a keresőoptimalizálást.

4 A keresőrendszerek jövője

Indexelés és visszakeresés

A téma kutatói rámutatnak, hogy a keresések gyengesége inkább az indexelendő dokumentumok indexelhetetlenségén, mint a keresőrendszere tehetségén múlik. Azt javasolják, hogy a mostani technikák finomítása helyett – ezek ugyanis homogén és jól strukturált dokumentumvilágról álmodnak – vegyék tudomásul, hogy az átlagkereső egy-két keresőszóval, semmi mással nem akar boldogulni. Néhány javasolt fejlesztési irány:

- intelligens ügynökszoftverek,
- természetes nyelvi indexelési módszerek,
- szabályalapú forrásfelderítés,
- képi és hangdokumentumok automatikus indexelése.

Felhasználói beállítások

Ha ismerik a felhasználók csoportjait, igényeiknek megfelelőbb kinézetű, szakterületükön használt paramétereket bekérő, csak a releváns adatbázisokban és csak az ő logikájuk szerint kereső, ezáltal hatékonyabb keresőrendszereket lehet kifejleszteni. Az általános célú keresők fel is ajánlanak ilyen felhasználói opciókat. Némely keresőrendszer felajánlja, hogy ha a találatok közül a felhasználó kiemelte a relevánsakat, akkor hasonlóság szerint automatikusan tovább keres, egyre jobb találatokért.

Összesítés

A keresőrendszerektől elvárható, hogy ne csak rendezve, de kategorizálva adják vissza a találatokat.

A keresés kiterjesztése, hatókör

Néhány keresőrendszer rendelkezésre bocsátja a találatok dokumentumaiban leggyakrabban előforduló szavak rangsorát, hogy a következő iteratív keresésben a felhasználó ezek közül egészíthesse ki kérdését.

A világhálón szaporodó multimédia adatformátumok nem hagyhatók figyelmen kívül. Ezek felismerése még nem tökéletes, az indexelésükhöz viszont már vannak jó ujjenyomatképző technikák.

Metaadatok

Ezek a bibliográfiai leírásra emlékeztető, helyettesítő dokumentumok az elsődleges adatbázisok felépülése során automatikusan előállhatnak, ha a dokumentumok kategorizálhatók, és a kategórián belül fegyelmezett formátumuk van. A keresés ilyenkor a metaadatokat tekinti végig. E módszer nemigen vihető át az általános keresőszolgáltatás világába, de tudományos, vagy más behatárolt felhasználói körben nagyon hatékony. Az általános keresők elé kapcsolt metakereső (Infoseek) párhuzamosan több keresőrendszer számára is konvertálja a kapott kérdést, az eredményeket pedig normalizálás után, összefésülve adja vissza.

Következtetések

Az információ-visszakeresés történelmi fordulópontján állunk. A mai számítógépekkel elfogadható időn belül óriási adatbázisokból meglehetősen bonyolult algoritmusokkal elég teljes és jó minőségű találati listát kaphatunk. A nagy adatbázisokat birtokló gazdálkodó szervezetek érdekeltek az információpiacon, támogatják a keresőrendszerek fejlesztését. Költségvetési pénzekből fejlődhet a digitális könyvtárak világa. A tudományos és üzleti szféra céljai közel kerültek egymáshoz.

5 A web keresőkről általában (keresőrobotok)

Keresőrobotoknak hívjuk (lásd még: kereső pókok, crawler, bot) azokat a programokat, melyek a Szabadszavas keresők számára gyűjtik össze a weblapokról az információkat (beindexelik az oldalakat), és bizonyos szabályok szerint automatikusan követik az egyes weboldalakon található linkeket. Ezek a robotok azonosítják magukat, és a rendes látogatottsági statisztika megmutatja, hogy mikor jártak utoljára oldalunkon, és hány oldalt látogattak meg.

A Szabadszavas keresők technológiának lényege, hogy lehetővé teszi a beindexelt információhalmazban (Pl. weboldalak, képek, stb.) a Kulcsszavak, Kereső-kifejezések alapján történő keresést. A szabadszavas keresők a Céltudatosan kereső felhasználók által messze leggyakrabban használt keresési formát jelentik.

A szabadszavas keresők minőségét elsősorban a Kereső adatbázisok mérete, az adatbázisok frissessége és a Találati rangsor Rangsorolásának minősége határozza meg. Emellett olyan szolgáltatásokkal is segíthetik a keresést, mint a Klasztizálás

Az internetes keresés másik gyakori formája a Tematikus katalógusok segítségével történő keresés. A tematikus keresőoldalak tartalma nem Keresőrobotok, hanem emberek által jön létre. Az ideális katalógus gazdái törekednek arra, hogy a katalógus minél hasznosabb oldalakat mutasson be, illetve, hogy besorolásuk lehetőleg minél pontosabb legyen.

A magyar téma szerinti keresőoldalak közül a legszervezettebb, legnagyobb és leglátogatottabb a startlap.hu konglomerátum. Leginkább itt érdemes keresni, és ide érdemes magyar weboldalt regisztrálni, ugyanis a többi magyar katalógus felhasználói tábor a kicsi, ezért csak elhanyagolható számú látogató érkezik ezekről az oldalakról. Említésre érdemes lehet még a vizsla24.hu linktár, linkcenter.hu, lapozz.hu webtár.

Az idegen nyelvű weboldalak jóval nagyobb nagyságrendje miatt ezek a katalógus oldalak nem annyira jól használhatóak, mivel majdnem hogy lehetetlenség ennyi honlapot kezelhető formában rendszerezni. A klasszikus katalógusok helyett/mellett ajánlott a del.icio.us és a Wikipédia használata az egyes témákban autentikus weblapok felleléséhez.

Metakeresőnek hívják az olyan keresőszolgáltatást, ami több kereső találatait összegzi; a Találati rangsort több kereső találatára alapján állítja össze, tehát nem saját adatbázisból dolgozik. (Hasonló hangzása ellenére semmi köze a Meta tagokhoz, tehát a Meta tagokat nem a Metakeresők miatt találták ki.) a metakeresőket akkor érdemes használni, ha szükség van az egyes metakeresők plusz szolgáltatásaira, mint például a Klaszterezés: a Clusty nevű metakereső a különböző keresőkből származó találatokat csoportokba is rendezi.

A Találati rangsor:

- Oldalunknak minél előbb kell szerepelnie a fontos Kulcsszavakra rákeresve, lehetőleg a Top 10 pozícióban
- A bekerülés mellett olyan részletnek kell megjelennie a keresőoldalon, ami meggyőzi a keresőt arról, hogy érdemes meglátogatni oldalunkat (Az oldal címét (TITLE) illetve a kulcsszavak szövegkörnyezetét jól kell megírni, és ezt ellenőrizni kell a Keresők segítségével történő nyomon követéssel.)
- Angol nyelvű szakirodalomban gyakran használják a SERP, SERPs? rövidítést (Search Engine Result Page, Search Engine Result Pages)

Látogatottsági statisztika: Mikor meglátogatunk egy oldalt, akkor a böngészőnk elküldi a szervernek szóló kérelem mellé, hogy milyen webcímről érkezett arra a bizonyos oldalra: Ez az ún. REFERRER, melyből többek között az is kinyerhető, hogy milyen kulcsszó beütése után jutott el oldalunkra látogatónk.

A referrer a böngészők által a webszervereknek elküldött információ, mely az adott kérelem elküldésekor aktuális URLt jelenti: ha például erről az oldalról a nyitólapra kattintasz, akkor a böngésző nemcsak azt küldi el a szervernek, hogy milyen lapot szeretnél látni, hanem elküldi annak a weblapnak is az URLjét, amit éppen most olvasol. A képek esetén hasonlóképpen azt az URLt küldi el, melybe az adott kép be van ágyazva. A refererek látogatottsági statisztika segítségével történő elemzésével rengeteg hasznos információt lehet szerezni egy weboldal látogatóiról.

Egyes böngészőkben, illetve egyes tűzfalak segítségével le lehet tiltani ezt a funkciót, továbbá szándékosan más URL-t is el lehet küldeni a valós referer helyett, de ezt referrer spam-nek hívják.

A robotok akár a keresőoldalakra regisztrálás révén, de leginkább a más webhelyekről az oldalunkra mutató külső hivatkozásokat követve érkeznek oldalunkra. A sok Teljes értékű link garancia egyrészt arra, hogy sokfajta kereső robotja látogassa az oldalt, másrészt pedig arra, hogy a gyakori látogatásokkal rendszeresen újraindexelje az oldal tartalmát.

A különböző Böngészők eltérő módon értelmezik az oldalon található HTML kódot, az esetleges hibákat és szabványtól eltérő megoldásokat is igyekeznek megjeleníteni, ezáltal a rossz kód sokszor fel sem tűnik az oldal alkotóinak, viszont a keresőrobotok nem feltétlenül rendelkeznek a böngészők hibátűrő képességével. Ez lehet akár szándékos is: egyrészt az értékes weboldalak valószínűleg jobban vannak megírva, mint a kevésbé értékesek; illetve a HTML értelmezésénél sokat számíthat a szükséges processzoridő, ezért feltehetően takarékosan bánnak vele (azaz nem biztos, hogy olyan nagy tudású, hibátűrő értelmezőket használnak, mint a böngészők).

Szintén az erőforrásokkal (pl. processzoridő) való takarékoság okán a keresőtechnológiák egyes, modern böngészők által értelmezhető technológiákat nem vesznek figyelembe (mint pl. JavaScript), ezáltal az ilyen, Problémás technológiák segítségével megírt oldalrészeket egyáltalán nem "látja" az adott kereső robotja.

A fenti okok miatt előfordulhat, hogy nem tudják az egész oldalt "elolvasni", vagy nem tudnak (nem akarnak) továbbmenni egy bizonyos oldalról.

Emiatt kell elkerülni a Problémás technológiákat és ragaszkodni a Webes szabványokhoz. Szöveges böngészők használatával, a weboldal forráskódjának átnézésével, illetve a Webes szabványok úgynevezett validátorainak segítségével meggyőződhetünk róla, hogy oldalunk nagy valószínűséggel problémamentesen indexszelhető-e. Mivel a keresőtechnológiák zöme titkos, ezért nem tudhatjuk, hogy az egyes robotok pontosan milyen módon értelmezve, mit látnak az (X)HTML, vagy az (X)HTML-hez hasonló, a Böngészőkben többbé-kevésbé jól megjelenő oldalakból, ezért a legtöbb, amit tehetünk, hogy ragaszkodunk a webes ajánlásokhoz. Ekkor van ugyanis a legnagyobb esélyünk arra, hogy a keresőpókok elé nem gördül semmilyen technikai akadály, és ez által webhelyünk egésze kereshető legyen.

Melyek a jó webstatisztika ismérvei?

A kereső-optimalizálás szempontjából különösen fontosak az alábbi adatok:

- milyen kulcsszavakra, Kereső-kifejezésekre keresve érkeztek oldalunkra,
- mely weboldalakon található Külső hivatkozásokról érkeztek oldalunkra,
- milyen keresők felől érkeznek a látogatók
- melyek a leglátogatottabb belépő oldalak (ahová leggyakrabban érkeznek látogatók más weboldalakról, vagy keresőoldalakról)
- melyek a leglátogatottabb oldalak általában
- milyen Keresőrobotok jártak az oldalon, mikor, és hány oldalt indexeltek

Bővebben a böngészőkről

Weboldalak megjelenítésére használt programok. A különböző böngészők másképpen értelmezik a Webes szabványokat, és még különfélebb módon reagálnak a hibás, nem szabványos kódokra.

Mozilla Firefox <http://mozilla.fsf.hu/>

Napjaink technikai értelemben véve egyik legjobb böngészője, mely a többi böngészőhöz képest számos előnyt nyújt (eltekintve a jól működő szabad forráskódú projektekre általában jellemző technológiai innovációtól és a biztonságosabb, stabliabb végterméktől), mint például:

- Az oldalon belüli keresést a leghatékonyabban oldja meg, mivel nem felugró ablakokba kell beírni a keresett kifejezést, ráadásul képes szöveges űrlapokban (textarea) való keresésre, ami nemcsak Wiki és Weblog szerkesztésnél nélkülözhetetlen.
- Rengeteg Firefox kiegészítő segítségével nemcsak az alapvető böngésző-funkciókat lehet bővíteni, hanem új funkciókkal is fel lehet ruházni a programot, mint pl. RSS olvasó, fájlcsereklő kliens.

Microsoft Internet Explorer

Az Internet Explorer-t elsősorban Windows operációs rendszerekre fejlesztették, de készült egy verzió Mac OS X operációs rendszerre is (5.2), mely azonban leginkább csak nevében hasonlít a Windowsos változathoz, és már hivatalosan nem is elérhető. A látogatók túlnyomó többsége a 6.0-os verziót használja, de kis százalékuk még mindig 5.5-ös vagy már a 7.0-ás verziókkal böngészik.

Az említett négyfajta verziónak négy, némileg különböző motorja van, ami máshogy tér el a szabványoktól, más hibái vannak, ezzel olykor igen körülményessé téve az összes verzióra való fejlesztést.

Az Internet Explorer-ek a Windows operációs rendszerek alapvető részét képezik, elterjedtségük ezzel is magyarázható.

A legújabb, nagymértékben átalakított kezelőfelületű és bővített funkciójú Internet Explorer 7, ami régebbi, sokak által használt operációs rendszerekre – mint pl. Windows 2000, Windows 98 –, nem elérhető. Ráadásul egy Windows rendszeren egyszerre csak egy Internet Explorer lehet fenn, ezért pl. a biztonsági frissítések módjára terjesztett IE 7-es böngészőt külön el kell távolítani, ha az egyelőre az elterjedtebb IE 6-tal szeretnénk megnézni oldalunk. Linux platformon létezik mindenesetre egy megoldás, mellyel egyszerre több Internet Explorer-t lehet futtatni egyszerre: <http://www.tatanka.com.br/ies4linux/>

Opera <http://www.opera.com>

Egy norvég cég által hosszú évek óta fejlesztett zárt forráskódú, de ingyenesen letölthető program. Előnye hogy gyors, számos hasznos kiegészítő funkciót nyújt, jól használható, illetve, hogy rengeteg operációs rendszeren fut. Persze nem minden operációs rendszerre érhető el mindig a legújabb verzió, hasonlóképpen a letölthető nyelvi verziók köre sem azonos, ezért lehet, hogy magyar nyelven csak egy régebbi verziót jelent még meg, vagy egyáltalán nem elérhető magyar nyelven az adott operációs rendszerre. A látogatóknak alig több mint egy százaléka használja ezt a böngészőt, a program pusztá fennmaradása nagyrészt innovatív funkcióinak és gyorsaságának köszönhető. A böngésző funkció mellett még mail-kliens is található a programban. Az Opera for Mobile pedig az okostelefonok PDA-k és egyéb mobiltelefonok széles választékára fejlesztett böngésző.

Konqueror és Safari

Az Unix/BSD/Linux operációs rendszerekre fejlesztett KDE grafikus környezet alapértelmezett böngészője. Az általa használt, KHTML-nek nevezett motor felhasználásával írták meg a Mac OS X alapértelmezett böngészőjét, a Safari-t. Gyakorlatilag elenyésző számú látogató használja ezeket a verziókat.

Netscape, Mozilla <http://mozilla.org>

Az internet hőskorának nagy klasszikus böngészői voltak a Netscape 4.x-es sorozat különböző verziói, mára már csak elenyésző számú felhasználó szörföl vele, akár csak az ezt követő Netscape verziókkal.

Az eredetileg a Netscape által fejlesztett böngészőmotor (gecko) nyílt forráskódúvá tételével jöttek létre a Mozilla, illetve a Firefox böngészők (Netscape 5.x). Mind a régi Netscape, mind a különböző Mozilla verziók gyakorlatilag majd' minden létező operációs rendszerre elérhetőek, mégis csak a felhasználók tizenvalahány százaléka használja, mivel nem alapértelmezett böngészője egyetlen operációs rendszernek sem. A gecko motort használó böngészők közül jelenleg a Firefox a legnépszerűbb, az eredeti Mozilla projekt pedig Seamonkey néven él tovább.

6 Egy intelligens web-kereső.

A fejezet célja, hogy bemutassa a már tárgyalt eszközöket egy intelligens web-kereső programon keresztül. Erre a célra a Copernic Agent Basic nevű programot választottam, melyre az Internet böngészése közben találtam rá.

A Web keresőmodulok nem adnak hozzáférést az Interneten rendelkezésre álló információ egészéhez. Továbbá mindegyik keresőmodul befedi a Web különböző területeit. Annyira ráadásul mikor miközben használjuk a legnagyobb keresőmodult, tárgyhoz tartozó információt szalasztunk el. A Copernic Agent Basic kezünkbe adja a képességet, hogy a Webből a lehető legtöbbet fedjenek be, a tárgyhoz tartozó találatok, és magas minőségi eredményeket kapjunk a több mint 90 keresőmodulból, amit kategóriákba csoportosítottak.

Egyetlen kérdésről a Copernic Agent Basic azonnal a többszörös keresőmodulok által ad jobb keresőmodul-eredményeket, miközben egyesülnek eredményeik.

A Copernic Agent működési leírása

Parallel search (Párhuzamos keresés):



A Copernic Agent-ben kiadott utasítás a keresésre, segítségül tud hívni több kereső motort párhuzamosan az internet átkutatására.



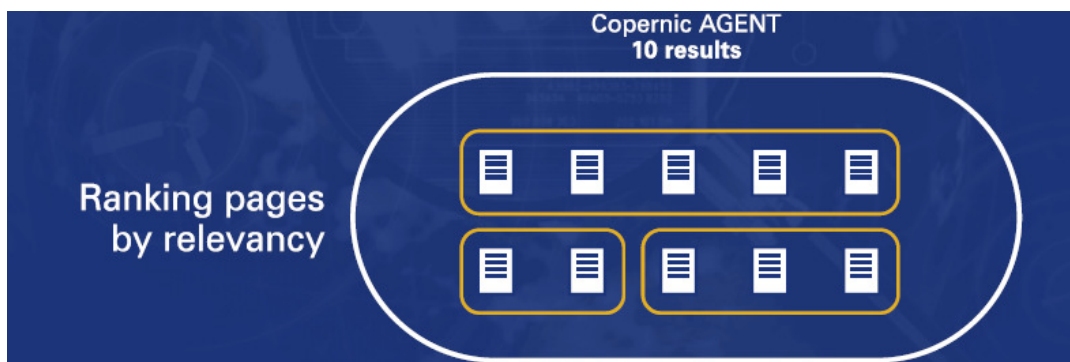
A keresés tárgyához tartozó weboldalak képezik a találatok alapjait, ezek a kereső eredmények:



Ezek után a találati listában szereplő oldalakból keletkezik egy egyesített lista, melyből, a program törli a duplikált találatokat:



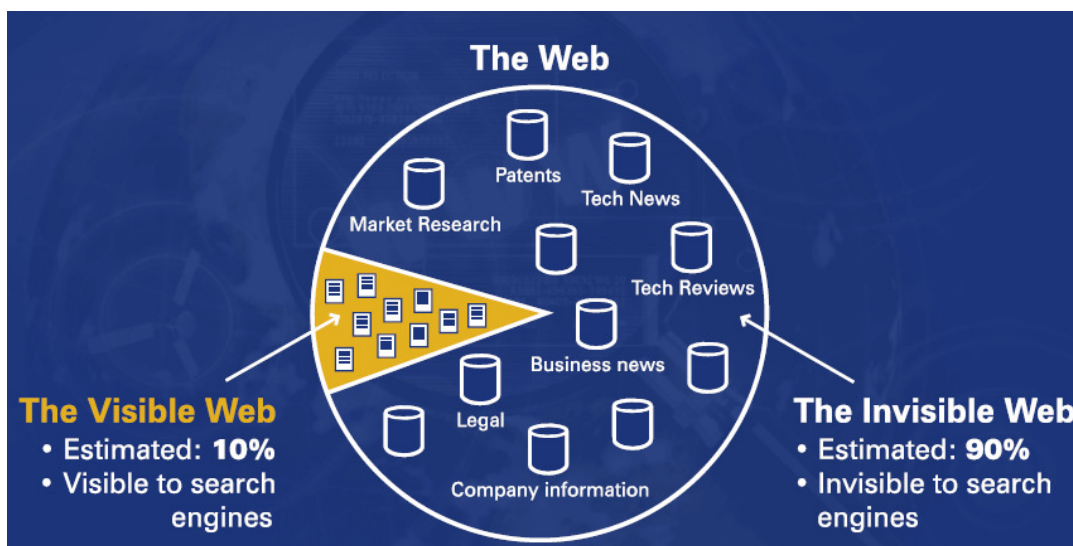
Az így elkészült eredményt aztán rangsorolja az oldalak fontossága szerint



Jobb web közvetítés, pontosabb tárgyhoz tartozóbb eredmények.

Láthatatlan világháló (Invisible Web)

A világháló 10%-átfedik be körülbelül rendszeresen a kereső motorok, ez a látható (Visible web), a többi a láthatatlan (Invisible web) része a világhálónak. A láthatatlan részt az üzleti hírek (Business news), szabadalmak (Patents), technikai hírek (Tech News), technikai áttekintések (Tech Reviews), törvényes (Legal), piackutatás (Market research), vállalati információ (Company information) teszik ki.



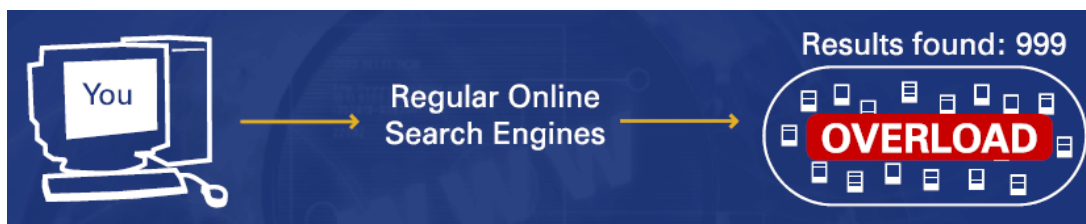
A Copernic Agent képes mind a két részterületben képes keresni bővítve ezzel az értékes találatok számát.



Segítségével hozzá lehet férni a „rejtett” információhoz, ezzel a keresésérdeklődés alapú lesz, a keresőkategóriákban.

Információ-többletteher (Information overload)

Míg a rendszeres online keresőmotorok találataik többletteherrel rendelkeznek, lehetséges többszörösen ismétlődő oldalakat találnak meg,



addig, a Copernic Agent az eredményeket elemzi,



a nem megfelelő kapcsolatokat törli, a lényegtelen eredményeket szűri,



végül, csoportosítja és rendezi eredményeket, az oldalakat és kiemelkedő kulcsfontosságú fogalmakat összefoglalja.

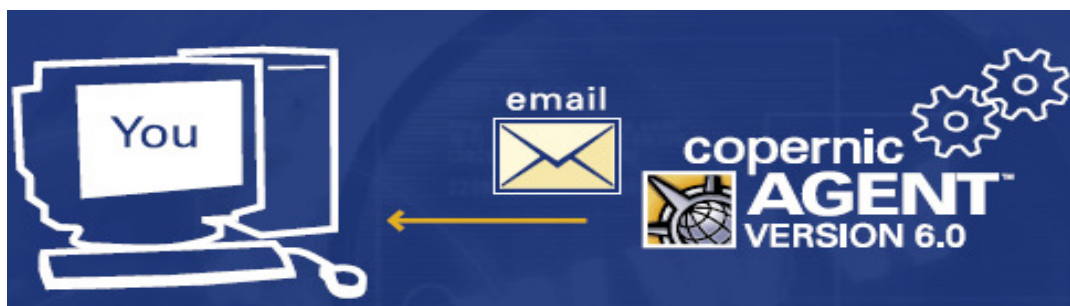


Ahol a keresőmodulok véget érnek ott kezdődik a Copernic Agent.



Több millió új oldal jelenik meg a világhálón mindennap. Nehéz elképzelni, hogy mi mindenről maradhatunk le! A Copernic Agent-tel lehetőség van arra, hogy értesítést kapjunk az azonos témájú új weboldalakról.

Segítségével naprakészek maradhatunk, vagy követhetjük új oldalak megjelenését, vagy e-mail jelentéseket kaphatunk, az új eredményekről automatikusan.



Oldalkövetés (Page tracking)

Hiábavaló ugyanazt az oldalt látogatni ezerszer a látszólagos változások miatt, a Copernic Agent automatikusan tudja felügyelni az oldalak tartalmát. Monitorozza a beállított weboldalt, és email-ben kihangsúlyozottan értesítést küld, az oldal változásairól.

A Copernic Agent jellemzői.

Keresés

A keresőmodulok és információforrások rendeződnek különböző csoportokba, azok, amik hasonlítanak, szakosodásuk szerint egy csoportba kerülnek.

A keresések tárgykörökbe csoportosulnak így megkönnyítve a keresést.

Segítségével, mintegy harmincféle kategória és téma szerint kereshetünk az interneten, kivonatot készíthetünk a weboldalakból, hogy könnyebben tudjunk dönteni, hogy a megfelelő oldalt találtuk-e meg.

Lehetőség van arra, hogy hozzáadjunk újabb keresőmodult és kategóriát.

Beállítható, testre szabható egy kedvenc kategóriákból álló csoport, melyen keresztül gyorsan elérhetővé válnak a leginkább használt keresőkategóriák.

Automatikusan ellenőrzi a törölt vagy nem megfelelő linkeket. Az oldalak nyelvezetét feltünteti, nem azonosan címzett, de tartalmilag azonos oldalakat észlelési, kiszűri. A lényegtelen találatokat automatikusan eltávolítja.

A felhasználónak lehetősége van arra, hogy logikai operátorok alkalmazásával (AND, OR, EXCEPT, NEAR), a weblap-tartalmát elemezze, ezzel finomítsa az eredményeket.

A már mentett kutatások eredményit frissíthetjük, melyeket kiemel a már meglévő találatoktól.

A keresési paramétereket módosíthatjuk, így lehetőség van a még pontosabb találatok elérésére.

A már meglévő keresési paraméterek másolhatók, annak érdekében, hogy gyorsan létre tudjak hozni egy hasonló keresést.

HTML-be vagy szövegformátumba könnyedén elektronikusan átpostázható.

Eredmények

A megtalált eredmények fontossági sorrendbe rendezettek, az azonos kapcsolatokkal rendelkező eredmények együttesen rendszerezettek.

A keresés eredményéről kiterjedt információt kaphatunk, mint például fontosságpontszámokról, weblap-kapcsolatokról, nyelvekről, felhasználói feljegyzésekről, dátumokról, keresőmodulokról.

Az eredménylistában feltűnő vizuális kiemeléseket használ, mint például az ikonok, aláhúzás, merész betűtípusok.

Az eredménylistákban és weblapokban található kulcsszavak kiemelve, melyekre kattintva böngészőnk megnyitja a hozzájuk kapcsolódó dokumentumokat.

Lehetőség van különböző csoportok valamilyen rendezési elv alapján történő rendezésre, fejlécre, kivonatra, címre, pontszámra, dátumra, keresésre, látogatási dátumra, vagy keresőmotorra.

Eredmények törlésére.

Web oldalkövetés

Előre beállítható időbeosztásokkal elérhető, hogy egy oldalt frissítsen többször egy nap egy napilapon, vagy heti esetleg havi lapon.

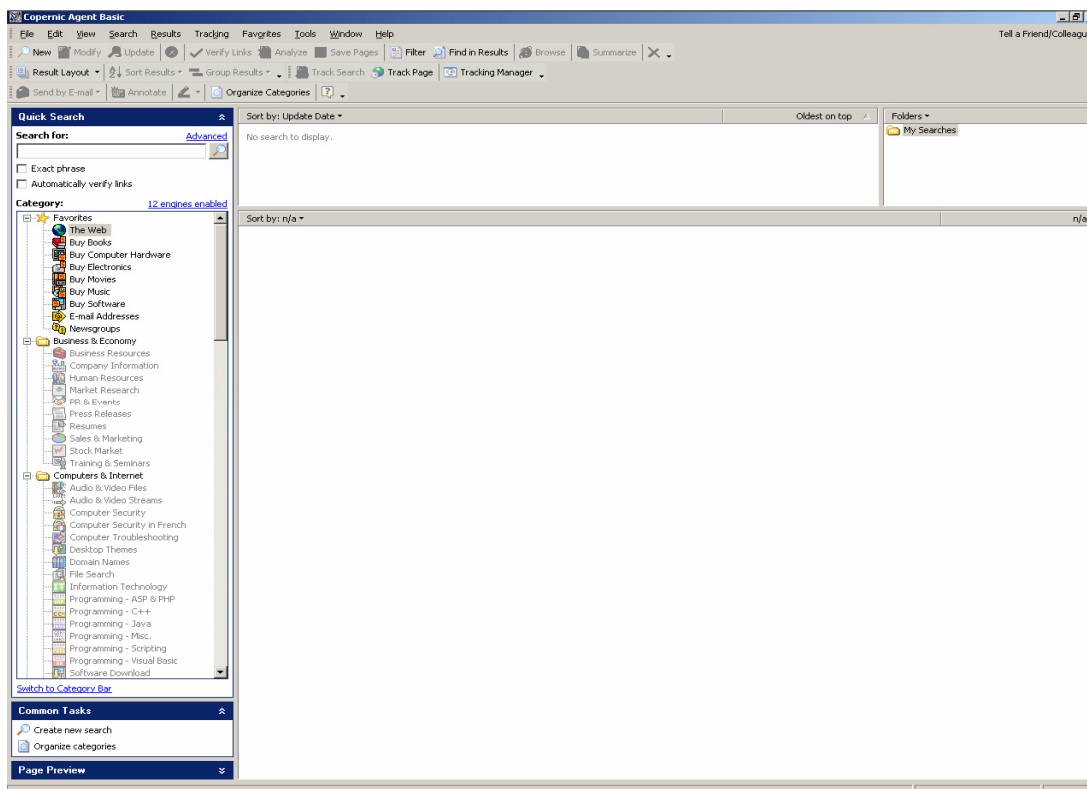
Egy követő feladatkezelő sok hasznos jellemzővel és parancsikkal.

Automatikus feladatteljesítési jelentés, amíg egy feladat fut, önműködő naplózással.

Egy ikon jelenik meg a Windows Taskbar-ban, amíg egy feladat folyamatban van, arra kattintva, megjelenik egy ablak a követés eredményével.

Keresés Indítása Copernic Agent Basic programban

A kereső rugalmas és könnyen kezelhető felületet biztosít, mely kiemelkedő keresőtapasztalatot és ergonomikusan tervezett, és ügyes interfészen keresztül tart kapcsolatot a felhasználóval.(1. ábra)



1. ábra A Copernic Agent Basic nyitó képernyője

A keresést egy kulcsszóra, egy adott témára kétféleképpen indíthatjuk. A Quick Search ablakban a Search for sorba írjuk a keresendő szót. Ha pl. névre keresünk, akkor érdemes idézőjelbe tenni: "Vezetéknév Keresztnév" hogy a teljes névre keressen.

A másik lehetőség, hogy a New gombra kattintunk, ekkor megjelenik a New Search ablak, és ide írjuk be a keresendő szót. Itt a "Modify search engine settings" linkre kattintva kijelölhetjük, melyik 11 keresőben keressen.

Egyszerre maximum 11 keresőben képes keresni a lehetséges 18-ból:

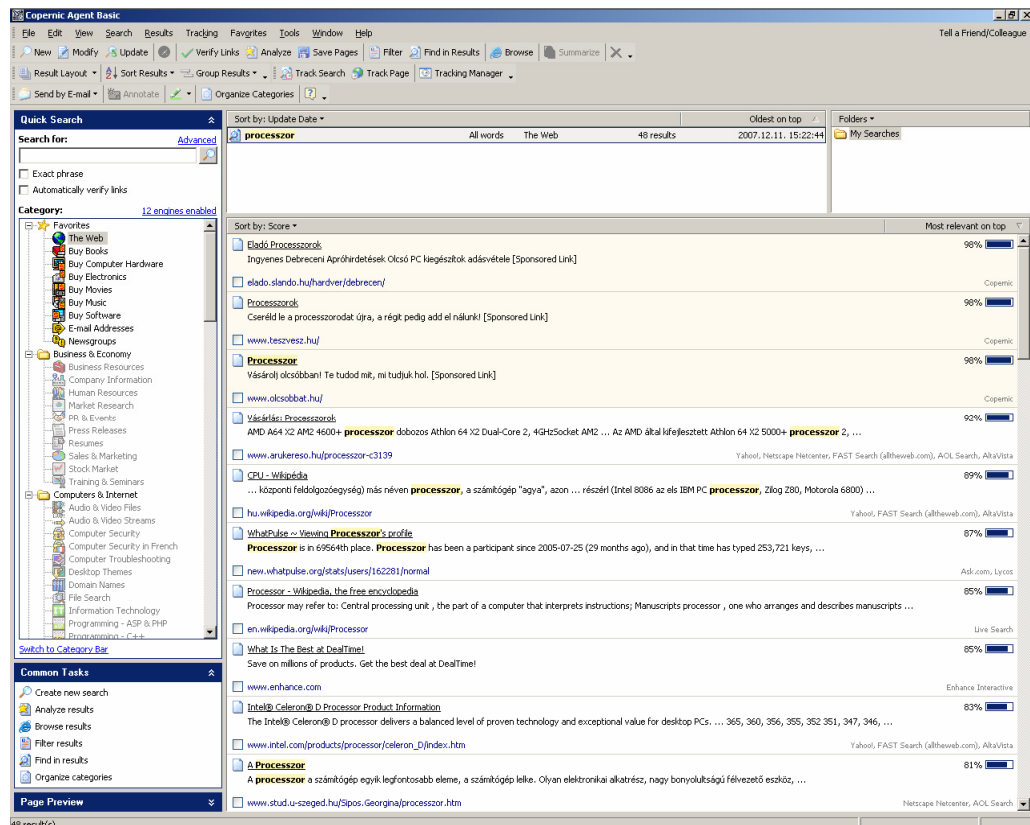
- AltaVista
- AOL.com search
- CompuServe
- Espotting
- EuroSeek
- FAST Search (alltheweb.com)
- FindWhat
- Google
- HotBot
- LookSmart
- Lycos
- Mamma.com
- MSN Web Search
- Netscape Netcenter
- Open Directory Project
- Teoma
- WiseNut
- Yahoo!

A webkereső modulok csak a keresőeredmények egy listáját hagyják hátra.

Copernic Agent megengedi, hogy sok mindent tegyünk a találatokkal, az eredményeken belül tudunk keresni, kapcsolat nélküli módban kereshetünk, menthetünk le oldalakat, rendezhetjük eredményit, e-mail-ben elküldhetők a keresőjelentések eredményei munkatársaknak, barátoknak.

A keresés eredményét elmenti, tehát később is meg tudjuk nézni, illetve frissíteni is tudjuk.

A keresés eredménye (Sort by: Score) egy összesítő ablakban jelenik meg, ahol a linkekre kattintva megnézhetjük az oldalakat.(2.ábra)



2. ábra A keresés eredménye

Amennyiben indítunk egy újabb keresést, abban az esetben letárolja, menti az előző állapotot és keletkezik egy újabb találati csoport, melyek között szabadon válthatunk.(3.ábra)

Ezek után lehetőségünk van alkalmazni az eddigiekben leírtak. A képen látható a találati ablakban (Sort by:) a rendezési lehetőségek, mely szabadon változtatható, továbbá a találatnak megfelelő százalékos arány, melyik keresővel találta az aktuális erdményt.

7 Konklúzió

A Copernic Agent egy Meta keresőmodul, láthatatlan internetes felfedező, online kutatósegéd és kiterjedt szerszámos doboz egy elegáns, könnyen kezelhető programban. Hasznos eszközökkel és jellemzőkkel szerelték föl, pontosan megtalálja, amit keresünk azáltal, hogy egyidejűleg jelentős keresőmodulokat és információforrásokat hív segítségül a Weben.

A keresőügynök jellemzői, hogy a kategóriák megfelelő keresőmodulokat céloznak meg; leáll információ-túlterhelés esetén, az eredményt analizálja, csoportosítja és összegzi, továbbá alkalmas naprakész eredmény és oldalkövetésre is. Teljes mértékben testre szabható, áttekinthető, rugalmasan kezelhető felülettel rendelkezik. Többek között egy gyors kereső, amely az integrált *Favorites* menü segítségével jelentős időt takarít meg, a kettős eredményeket kiküszöböli, a törött kapcsolatokat eltávolítja és igazolja őket, eredménylistákban és weblapokban kihangsúlyozza a kulcsszavakat.

A kapcsolat nélküli böngészésért oldalakat ment le, és az eredményekben levő kulcsszavak vagy weblapok boolean algebrai operátorok segítségével leszűkíthetők.

A Copernic Agent egyedi dossziékkal, egy részletes keresőtörténelmet tart nyilván, azért hogy később az eredményeket visszakereshessük és megnézhessük őket annak érdekében, hogy könnyedén új eredményeket találhassunk meg. A sok értékes jellemzője között ez a szoftver több lehetőséget ajánl fel arra, hogy menüket, eszköztárakat, keresőlistákat, keresőoperációkat és opciókat szabhassunk testre. Szintén jellemzi a programot, hogy automatikus szoftver és motor frissítő. A Copernic Agent különösen Windowsba való kiterjesztésekre specializálódott, ami integrál egy olyan sok funkcióból álló eszköztárt az Internet Explorerbe, ami megengedi nekünk, hogy kulcsszavakat nyomozzunk ki olyan weblapokban, amiket használunk, és új kereséseket hozhatunk létre.

A Copernic Agent egy tágra nyitott olló, ami a párhuzamos keresések segítségével sokkal bővebben kutatja át az Internetet. Eléri a rejtett információkat, ami a láthatatlan világhálón található, a kereséseink könnyebbek, gyorsabbak és hatékonyabbak lesznek, mint valaha.

8 Köszönetnyilvánítás

Szeretnék köszönetet mondani Mecsei Zoltánnak, hogy tanácsaival, javaslataival segítette a szakdolgozatom elkészítését.

9 Irodalmi jegyzék

- <http://www.copernic.com/>
- http://tmt.omikk.bme.hu/issue.html?issue_id=18
- <http://weblabor.hu/cikkek/googlealapjai1>
- http://webni.innen.hu/B_c3_b6ng_c3_a9sz_c5_91k
- <http://www.tutorial.hu/alapveto-seo-tanacsok/>
- www.geo.u-szeged.hu/~laci/FIR-T%E9rinformatiak%20alapjai/FIR-X.PPT