



دانشگاه علوم پزشکی تهران

معاونت توسعه مدیریت و برنامه ریزی منابع

آموزش الکترونیکی ضمن خدمت کارکنان



جزوه کمک آموزشی



نام دوره آموزشی : جستجو در موتورهای جستجو

ساعت دوره آموزشی : ۱۶ ساعت



فهرست

۱. اهداف رفتاری ۳
۲. مقدمه ۴
۳. موتورهای جستجو و علت پیدایش آنها ۴
۴. جایگاه موتورهای جستجو ۶
۵. روش کار موتور جستجو ۸
 - عنکبوت
 - خزنده
 - پایگانی کننده
 - پایگاه داده
 - سیستم رتبه بندی
۶. دسته بندی موتورهای جستجو و نحوه کار آنها ۱۳
 - موتورهای جستجوی پیمایشی
 - فهرستهای تکمیل دستی
 - موتورهای جستجوی ترکیبی با نتایج مختلط
 - بررسی یک موتور جستجوی پیمایشی
 - موتورهای جستجو چه مقداری از کل صفحات وب را برای پیدا کردن پاسخ، جستجو می کنند؟
 - علت تفاوت نتایج حاصل از موتورهای جستجوی گوناگون
 - مقایسه کارایی چند موتور جستجو کاربردی
 - مقایسه برخی از ویژگیهای چند موتور جستجو کاربردی
 - راهکارهایی برای کوتاه کردن مسیر جستجو
 - مهم ترین ترفندهای گوگل



۷. بررسی کمی و کیفی موتورهای جستجو ۲۴
- اندازه گزارش شده
 - عمق صفحات
 - چگونه موتورهای جستجو صفحات وب را رتبه بندی می کنند؟
۸. روش کار موتورهای جستجو در رتبه بندی سایتها ۲۸
- بهینه سازی رتبه وبسایت در موتورهای جستجو
 - فاکتورهای کاهش رتبه در موتورهای جستجو
۹. سرویس جستجوی دایرکتوری ها ۳۴
- تفاوت موتورهای جستجو با دایرکتوری ها (فهرست ها)
۱۰. خودآزمایی ۳۶
۱۱. پاسخنامه ۳۶
۱۲. کلید واژگان ۳۷
۱۳. منابع ۳۹



۱. اهداف رفتاری

در پایان دوره از فراگیران انتظار می‌رود:

۱. روش کار موتورهای جستجو را توضیح دهند.
۲. روش‌ها و ترفندهای جستجو از طریق اینترنت را شرح دهند.
۳. روش کار موتورهای جستجو در رتبه بندی سایت‌ها را بیان کنند.
۴. طریقه بهینه‌سازی رتبه وب سایت‌ها در موتورهای جستجو را توضیح دهند.

۲. مقدمه

از سال ۱۹۹۲ یک منبع بسیار مهم دیگر به منابع اطلاعاتی اضافه شد و آن شبکه جهانی وب (World Wide Web) یا همان اینترنت است. به جز کاربرد در زمینه اطلاع رسانی اینترنت کاربردهای بی شمار و غیرقابل تصویری در تجارت و سایر زمینه های زندگی نیز پیدا کرده است. به بیان دیگر امروزه در هر زمینه ای که بتوان تصور کرد، اینترنت یک ابزار مؤثر و توانمند به حساب می آید.

موتورهای جستجو ابزارهایی برای پیدا کردن اطلاعات بر روی شبکه اینترنت هستند. این ابزارها از زمانی که شبکه اینترنت به عنوان شاهراه و منبع اطلاعاتی مطرح شد بوجود آمدند تا به کاربران در جستجوی اطلاعات مورد نظر در میان انبوه اطلاعات موجود در شبکه کمک کنند. در حالی که میزان اطلاعات در شبکه به شدت افزایش می یابد و هم اکنون از مرز ۸۰۰ میلیون صفحه گذشته است. استفاده از موتورهای جستجو برای هر کاربر، امری اجتناب ناپذیر می نماید. بیشتر مردم به ابزارهایی علاقمندند که اطلاعات را در شبکه برایشان پیدا کند. اصولاً دو نوع سرویس جستجو در وب وجود دارد :

➤ موتورها (Engines)

➤ دایرکتوریها (Directories)

که در ادامه به توضیح هر یک خواهیم پرداخت.

۳. موتورهای جستجو و علت پیدایش آنها

با وجود حجم روزافزون طراحی و راه اندازی سایت های وب مختلف، نیاز به دستیابی به مرکزی برای شناسایی این پایگاه ها برای استفاده بهتر و بیشتر کاربران وب، امر مهمی به شمار آمد. از آنجاکه سایت ها و مطالبی که در وب منتشر می شوند، توسط هیچ مرکز رسمی بین المللی مسئول در اینترنت اعلام نمی شود، تنها راه پیدا کردن یک موضوع، اطلاع داشتن دقیق از آدرس آن سایت به نظر می رسد.

موتورهای جستجو برای سهولت دسترسی کاربران به مطالب موجود در سایت های وب راه اندازی شده اند. در واقع بدون استفاده از موتورهای جستجو قادر نخواهیم بود تا از به روزرسانی های سایت های وب اطلاع کسب کنیم و نیازهای تحقیقاتی و آموزشی و تجاری و خبری و ... خود را پوشش دهیم.

مهم ترین توانایی موتورهای جستجو جذب مخاطبان واقعی سایت است. مخاطبان وبگرد و سرگردان معمولاً ارزش چندانی ندارند.

امروزه بر روی اینترنت و مهمترین سرویس آن (وب)، صدها میلیون صفحه حاوی اطلاعات وجود دارد. کاربران اینترنت با آگاهی از آدرس یک سایت قادر به اتصال به سایت موردنظر و استفاده از منابع اطلاعاتی موجود بر روی سایت خواهند بود. ما با دریایی از اطلاعات مواجه هستیم. در صورتی که قصد یافتن اطلاعاتی خاص را داشته باشیم، از چه امکاناتی در این زمینه می توان استفاده کرد؟ برای جستجو و یافتن اطلاعات مورد نیاز از مراکز جستجوی اطلاعات در اینترنت استفاده می-گردد. به مراکز فوق Search engines نیز می گویند.

مراکز جستجو در اینترنت، نوع خاصی از سایت های موجود در وب بوده که با هدف کمک برای یافتن اطلاعات، ایجاد شده اند. مراکز جستجو در اینترنت بمنظور پاسخگویی به کاربران متقاضی و جستجوکنندگان اطلاعات از سه روش متفاوت استفاده می نمایند. نحوه عملکرد سه روش بایکدیگر مشابه بوده و تنها تفاوت موجود میدان انتخاب شده برای عملیات جستجو است.

✚ اینترنت و یا بخشی از آن براساس کلمات مهم جستجو می گردد.

✚ از کلمات پیدا شده یک ایندکس به همراه محل مربوط به هریک ایجاد می نمایند.

✚ به کاربران امکان جستجو برای کلمات خاص و یا ترکیبی از آنها که در فایل ایندکس موجود می باشد، داده می شود.

مراکز جستجوی اولیه در اینترنت، صرفاً اطلاعات مربوط به چندین هزار صفحه وب را ایندکس و روزانه دو تا سه هزار کاربر متقاضی به آنها مراجعه می کردند. مراکز جستجوی فعلی اولیه در اینترنت اطلاعات مربوط به صدها میلیون صفحه را ایندکس نموده و روزانه به بیش از دهها میلیون متقاضی پاسخ می دهند.

موتورهای جستجو موتورهای ربات هایی هستند که صفحات وب (Web pages) را می پیمایند و صفحات جدید را پیدا می کنند. این ربات ها صفحات وب را می خوانند و تمام یا قسمتی از متن را در یک بانک اطلاعاتی یا فهرستی که شما امکان دستیابی به آن را داشته باشید قرار می دهند هیچکدام از آنها تمام شبکه را پوشش نمی دهند ولی بعضی از آنها واقعاً بزرگ هستند.

موتورهای جستجو اصلی در این زمینه عبارتند از: Alta vista, Google, Fast All the Web, Askjeeves, Inktomi .

"Inktomi" در اصل یک سایت جستجو نیست ولی MSN, Hot bot را تغذیه اطلاعاتی می کند. "Fast" بیشتر پرتالهایی

مانند "lycos" را تغذیه می کند .

موتورهای جستجو باید اولین انتخاب شما باشد، زمانیکه دقیقاً می دانید که دنبال چه چیزی می گردید. آنها همچنین

قسمت خیلی بزرگتری از وب را در مقایسه با دایرکتوریها می پوشانند.



هرچند فرق میان دایرکتوریها و موتورهای جستجو آنقدر که در گذشته مشخص بود، اکنون نیست همه دایرکتوریهای جستجو نتایج جستجو را ابتدا از موتورهای جستجو پیدا می کنند و اگر در آن جا جستجو موفق نبود، به سراغ دایکتوری خودشان می روند مثلاً "Yahoo" از موتور جستجوی "Google" برای این منظور استفاده می کند .

از طرف دیگر بعضی موتورهای جستجو اطلاعات را از دایرکتوریهای جستجو تهیه می کنند، قبل از اینکه از بانک اطلاعاتی موتورهای جستجو اطلاعاتی را به شما بدهند.

Ask Jeeves سعی می کند جوابها را به سئوالات شما از زبان طبیعی بانک اطلاعاتی خودش بیرون بکشد. یعنی یک سری پاسخها به پرسشهای متداول را در خود دارد.

Lycos, Google, AOL به شما قابلیت دسترسی به دایرکتوریهای باز را می دهند. Alta Vista, MSN یک سری سلسله دایرکتوری دارند مشابه Yahoo . اینها براساس Look smart directory ساخته شده اند. اگر آنچه را که جستجو می کنید نتوانستید پیدا کنید، نتایج را در سایتهای مشابه برای شما پیدا خواهد کرد.

۴. جایگاه موتورهای جستجو

موتورهای جستجو (search engines) اکنون در فعالیتهای اینترنتی، نقش غیرقابل انکاری پیدا کرده اند. بسیاری از تحقیقات نشان داده اند که موتورهای جستجو ابزاری مهم و محبوب برای یافتن اطلاعات مورد نیاز کاربران است. در آگوست ۲۰۰۴، مطالعه ای توسط مؤسسه Pew و ComScore در زمینه جایگاه موتورهای جستجو انجام شده است که خلاصه ای از نتایج آن را در این نوشته می خوانید. این تحقیق، با استفاده از پرسش تلفنی از ۱۳۹۹ کاربر اینترنتی در سراسر آمریکا انجام گرفته است.

مطالعه مورد اشاره نشان داده است که استفاده از موتورهای جستجو یکی از محبوب ترین فعالیتهای اینترنتی است. اکثر اوقات، استفاده از موتورهای جستجو، پس از پست الکترونیک در جایگاه دوم قرار دارد. البته گاهی که اتفاق مهمی رخ می دهد (مثل جنگ در عراق) دریافت اخبار از طریق اینترنت بر استفاده از موتورهای جستجو پیشی می گیرد. از نظر آماری ۸۴ درصد کاربران اینترنتی بیان داشته اند که از موتورهای جستجو استفاده می کنند. حجم جستجو نیز قابل توجه است. هر کاربر اینترنتی به طور میانگین ۳۳ جستجو انجام می دهد. به این ترتیب، فقط در آمریکا، با استفاده از ۲۵ موتور جستجوی متداول، ۳/۹ میلیارد جستجو در ماه انجام می شود. همچنین در حالی که ۶۴ میلیون نفر از شهروندان آمریکا هر روز به اینترنت متصل می شوند، بیش از ۳۸ میلیون نفر از آنان از موتورهای جستجو استفاده می کنند.

در مورد میزان سودمندی موتورهای جستجو نیز بیشتر کاربران، اعتماد و رضایت خود را هنگام کار با موتورهای جستجو نشان داده‌اند. ۸۷ درصد کاربران ابراز کرده‌اند که بیشتر اوقات به آنچه جستجو کرده‌اند، دست یافته‌اند. از این- گروه ۲۰ درصد همیشه از نتایج جستجو رضایت داشته‌اند.

همچنین بسیاری از کاربران مهارت خود را در استفاده از موتورهای جستجو افزایش داده‌اند. ۹۲ درصد جستجوگران اعلام کرده‌اند که به مهارت خود در یافتن نتایج اطمینان دارند. بیش از نیمی از آنان گفته‌اند که کاملاً مطمئن هستند و می‌توانند آنچه را در نظر دارند با استفاده از جستجوی اینترنتی پیدا کنند.

در دسترس بودن موتورهای جستجوی مطمئن و ساده، نحوه دستیابی کاربران به اطلاعات را تغییر داده است. بسیاری از کاربران به میزان زیادی، برای دستیابی به اطلاعاتی که برای آنان بسیار حیاتی است، وابسته به موتورهای جستجو هستند: ۴۴ درصد جستجوگران بیان داشته‌اند اطلاعاتی که از طریق موتورهای جستجو می‌یابند، برایشان کاملاً ضروری است. اهمیت این آمار آنگاه دوچندان می‌شود که بدانیم یک سوم جستجوکنندگان از طریق اینترنت بیان کرده‌اند که بدون موتورهای جستجو نمی‌توانند به فعالیت‌های خود ادامه دهند. اهمیت موتورهای جستجو برای نیمی از کاربران کمتر است. آنان ابراز داشته‌اند که موتورهای جستجو علاقه دارند، ولی در صورت لزوم می‌توانند اطلاعات مورد نظر خود را از روش‌های دیگر به دست آورند. در میان موتورهای جستجوی متداول میزان رضایت از گوگل (Google) بسیار زیاد است. در این مطالعه، ۴۷ درصد پرسش‌شوندگان اعلام کرده‌اند که گوگل انتخاب اول آنان برای جستجو است.

یاهو (Yahoo) با ۲۶ درصد در مقام دوم قرار دارد. پس از این دو، AOL، MSN، Askjeeves و Netscape با ۷، ۵، ۲ و ۱ درصد قرار گرفته‌اند. به این ترتیب نتایج این مطالعه برای گوگل و یاهو بسیار خوشحال‌کننده است. میزان اعتماد کاربران به موتورهای جستجو نیز جالب است. ۴۴ درصد کاربران گفته‌اند تنها از یک موتور جستجو برای یافتن نتایج مورد نظرشان استفاده می‌کنند. ۴۸ درصد نیز ۲ یا ۳ موتور جستجو را تجربه می‌کنند.

به عنوان بخشی از این مطالعه، موسسه ComScore برای سنجش میزان رضایت کاربران از آنان سؤال کرد کدام ویژگی موتورهای جستجو برای آنان بیشتر مهم بوده است و موتور جستجوی مورد استفاده آنان از این نظر در چه جایگاهی قرار دارد.

نتیجه این بخش از مطالعه آن است که؛ گرچه میزان رضایت کاربران از موتورهای جستجو به طور کلی بسیار زیاد است، ولی فاصله نسبتاً زیادی میان توقع آنان از نظر ویژگی‌های مورد نظرشان با آنچه از موتورهای جستجو کسب کرده‌اند وجود دارد. برای مثال، ۹۱ درصد از پرسش‌شوندگان، عبارت «نتایجی برمی‌گرداند که بانایزهای شما تطبیق دارد»، را به عنوان



مهمترین ویژگی مدنظر اعلام کرده‌اند، درحالی که تنها ۶۶ درصد آنان گفته‌اند که موتور جستجوی مورد علاقه‌شان توقع آنان را از این نظر برآورده می‌سازد.

همچنین ۸۵ درصد کاربران محفوظ نگه داشتن اطلاعات مورد جستجو را مهم دانسته‌اند، اما تنها ۵۵ درصد اعلام کرده‌اند که موتور جستجوی مورد علاقه‌شان این موضوع را رعایت می‌کند. از دیگر پارامترهای مهم مورد نظر کاربران می‌توان به - این موارد اشاره کرد: ۹۰ درصد کاربران، راحتی استفاده را عامل مهمی می‌دانند. از نظر ۸۹ درصد آنان، سرعت بازگرداندن - اطلاعات بسیار مهم است.

جالب است که پیشنهاد کلمه جستجوی جدید هنگامی که کلمه خوبی از سوی کاربر جستجو نشده است، از ویژگی‌های مورد علاقه ۷۰ درصد کاربران بوده است.

۵. روش کار موتور جستجو

وقتی جستجویی در یک موتور جستجوگر انجام و نتایج جستجو ارائه می‌شود، کاربران در واقع نتیجه کار بخش‌های متفاوت موتور جستجوگر را می‌بینند. موتور جستجوگر قبلاً پایگاه داده‌اش را آماده کرده است و این گونه نیست که درست در همان - لحظه جستجو، تمام وب را بگردد. بسیاری از خود می‌پرسند که چگونه امکان دارد گوگل در کمترین زمان یک ثانیه تمام سایت‌های وب را بگردد و میلیون‌ها صفحه را در نتایج جستجوی خود ارائه کند؟
نه گوگل و نه هیچ موتور جستجوگر دیگری توانایی انجام این کار را ندارند. همه آنها در زمان پاسخ‌گویی به کاربران، تنها در پایگاه داده‌ای که در اختیار دارند به جستجو می‌پردازند و نه در وب!
موتور جستجوگر به کمک بخش‌های متفاوت خود، اطلاعات مورد نیاز را قبلاً جمع‌آوری، تجزیه و تحلیل می‌کند و آنرا در پایگاه داده‌اش ذخیره می‌نماید و به‌هنگام جستجوی کاربر تنها در همین پایگاه داده می‌گردد.

بخش‌های مجزای یک موتور جستجوگر عبارتند از:

Spider یا عنکبوت

Crawler یا خزنده

Indexer یا بایگانی کننده

Database یا پایگاه داده

Ranker یا سیستم رتبه بندی

الف - عنكبوت

عنكبوت (Spider) یا روبات (Robot)، نرم افزاری است که کار جمع آوری اطلاعات مورد نیاز یک موتور جستجوگر را برعهده دارد. اسپایدر به صفحات مختلف سر می زند، محتوای آنها را می خواند، اطلاعات مورد نیاز را جمع آوری می کند و آنرا در اختیار سایر بخش های موتور جستجوگر قرار می دهد. کار یک اسپایدر، بسیار شبیه کار کاربران وب است. همانطور که کاربران، صفحات مختلف را بازدید می کنند، اسپایدر هم درست این کار را انجام می دهد با این تفاوت که اسپایدر کدهای HTML صفحات را می بیند اما کاربران نتیجه حاصل از کنار هم قرار گرفتن این کدها را می بینند.

اما یک اسپایدر آنرا چگونه می بیند؟

برای این که شما هم بتوانید دنیای وب را از دیدگاه یک اسپایدر ببینید، کافی است که کدهای HTML صفحات را مشاهده کنید.

آیا این دنیای متنی برای شما جذاب است؟

اسپایدر، به هنگام مشاهده صفحات، از خود بر روی سرورها رد پا برجای می گذارد. شما اگر اجازه دسترسی به آمار دید و بازدیدهای صورت گرفته از یک سایت و اتفاقات انجام شده در آنرا داشته باشید، می توانید مشخص کنید که اسپایدر کدام یک از موتورهای جستجوگر صفحات سایت را مورد بازدید قرار داده اند. یکی از فعالیتهای اصلی که در SEM انجام می شود، تحلیل آمار همین دید و بازدیدها می باشد.

اسپایدرها کاربردهای دیگری نیز دارند، به عنوان مثال عده ای از آنها به سایت های مختلف مراجعه می کنند و فقط به بررسی فعال بودن لینک های آنها می پردازند و یا به دنبال آدرس پست الکترونیکی (Email) می گردند.

ب - خزنده

خزنده (Crawler)، نرم افزاری است که به عنوان یک فرمانده برای اسپایدر عمل می کند. آن مشخص می کند که اسپایدر کدام صفحات را مورد بازدید قرار دهد. در واقع کراولر تصمیم می گیرد که کدام یک از لینک های صفحه ای که اسپایدر در حال حاضر در آن قرار دارد، دنبال شود. ممکن است همه آنها را دنبال کند، بعضی ها را دنبال کند و یا هیچ کدام را دنبال نکند. کراولر، ممکن است قبلاً برنامه ریزی شده باشد که آدرس های خاصی را طبق برنامه، در اختیار اسپایدر قرار دهد تا از آنها دیدن کند. دنبال کردن لینک های یک صفحه به این بستگی دارد که موتور جستجوگر چه حجمی از اطلاعات یک سایت را می تواند در پایگاه داده اش ذخیره کند و همچنین ممکن است اجازه دسترسی به بعضی از صفحات به موتورهای جستجوگر داده نشده باشد.

شما به عنوان دارنده سایت، همان طور که دوست دارید موتورهای جستجوگر اطلاعات سایت شما را با خود ببرند، می توانید آنها را از بعضی از صفحات سایت تان دور کنید و اجازه دسترسی به محتوای آن صفحات را به آنها ندهید. تنظیم میزان دسترسی موتورهای جستجوگر به محتوای یک سایت توسط پروتکل Robots انجام می شود که در مقالات دیگر سایت به آن پرداخته شده است. به عمل کراولر، خزش (Crawling) می گویند.

ج - پایگانی کننده

تمام اطلاعات جمع آوری شده توسط اسپایدر در اختیار ایندکسر (Indexer) قرار می گیرد. در این بخش اطلاعات ارسالی مورد تجزیه و تحلیل قرار می گیرند و به بخش های متفاوتی تقسیم می شوند. تجزیه و تحلیل بدین معنی است که مشخص می شود اطلاعات از کدام صفحه ارسال شده است، چه حجمی دارد، کلمات موجود در آن کدام است، کلمات چندبار تکرار شده است، کلمات در کجای صفحه قرار دارند و ...

در حقیقت ایندکسر، صفحه را به پارامترهای آن خرد می کند و تمام این پارامترها را به یک مقیاس عددی تبدیل می کند تا سیستم رتبه بندی بتواند پارامترهای صفحات مختلف را باهم مقایسه کند. در زمان تجزیه و تحلیل اطلاعات، ایندکسر برای کاهش حجم داده ها از بعضی کلمات که بسیار رایج هستند صرف نظر می کند. کلماتی نظیر a, an, the, www, is و ... از این گونه کلمات هستند.

د - پایگاه داده

تمام داده های تجزیه و تحلیل شده در ایندکسر، به پایگاه داده (Database) ارسال می گردد. در این بخش داده ها گروه بندی، کدگذاری و ذخیره می شود. همچنین داده ها قبل از آنکه ذخیره شوند، طبق تکنیکهای خاصی فشرده می شوند تا حجم کمی از پایگاه داده را اشغال کنند.

یک موتور جستجوگر باید پایگاه داده عظیمی داشته باشد و به طور مداوم حجم محتوای آنرا گسترش دهد و البته اطلاعات قدیمی را هم به روزرسانی نماید. بزرگی و به روز بودن پایگاه داده یک موتور جستجوگر برای آن امتیاز محسوب می گردد. یکی از تفاوت های اصلی موتورهای جستجوگر در حجم پایگاه داده آنها و همچنین روش ذخیره سازی داده ها در پایگاه داده است.

و - سیستم رتبه بندی

بعد از آنکه تمام مراحل قبل انجام شد، موتور جستجوگر آماده پاسخ گویی به سوالات کاربران است. کاربران چند کلمه را در جعبه جستجو (Search Box) آن وارد می کنند و سپس با فشردن Enter منتظر پاسخ می مانند. برای پاسخ گویی به

درخواست کاربر، ابتدا تمام صفحات موجود در پایگاه داده که به موضوع جستجو شده، مرتبط هستند، مشخص می‌شوند. پس از آن سیستم رتبه‌بندی (Ranker) وارد عمل شده، آنها را از بیشترین ارتباط تا کمترین ارتباط مرتب می‌کند و به عنوان نتایج جستجو به کاربر نمایش می‌دهد.

حتی اگر موتور جستجوگر بهترین و کامل‌ترین پایگاه داده را داشته باشد اما نتواند پاسخ‌های مرتبطی را ارائه کند، یک موتور جستجوگر ضعیف خواهد بود. در حقیقت سیستم رتبه‌بندی قلب تپنده یک موتور جستجوگر است و تفاوت اصلی موتورهای جستجوگر در این بخش قرار دارد.

سیستم رتبه‌بندی برای پاسخ‌گویی به سوالات کاربران، پارامترهای بسیاری را در نظر می‌گیرد تا بتواند بهترین پاسخ‌ها را در اختیار آنها قرار دهد. حرفه‌ای‌های دنیای SEM به طور خلاصه از آن به Algo (الگوریتم) یاد می‌کنند.

الگوریتم، مجموعه‌ای از دستورالعمل‌ها است که موتور جستجوگر با اعمال آنها بر پارامترهای صفحات موجود در پایگاه داده‌اش، تصمیم می‌گیرد که صفحات مرتبط را چگونه در نتایج جستجو مرتب کند. در حال حاضر قدرتمندترین سیستم رتبه‌بندی را گوگل در اختیار دارد.

می‌توان با ادغام کردن اسپایدر با کراولر و همچنین ایندکسر با پایگاه داده، موتور جستجوگر را شامل سه بخش زیر دانست که این‌گونه تقسیم‌بندی هم درست می‌باشد:

کراولر

بایگانی

سیستم رتبه‌بندی

تذکر: برای سهولت در بیان مطالب بعدی، هرگاه صحبت از بایگانی کردن (شدن) به میان می‌آید، مقصود این است که صفحه تجزیه و تحلیل شده و به پایگاه داده موتور جستجوگر وارد می‌شود.

مثالی از روش کار موتور جستجو

برای آنکه تصور درستی از نحوه کار یک موتور جستجوگر داشته باشید داستان نامتعارف زیر را با هم بررسی می‌کنیم.

داستان ما یک شکارچی دارد. او تصمیم به شکار می‌گیرد:

- **کار کراولر:** او قصد دارد برای شکار به منطقه حفاظت شده ابیورد، واقع در شهرستان درگز (شمالی‌ترین شهر خراسان بزرگ) برود.

- **پروتکل Robots:** ابتدا تمام محدودیت‌های موجود برای شکار در این منطقه را بررسی می‌کند:

✚ آیا در این منطقه می توان به شکار پرداخت؟

✚ کدام حیوانات را می توان شکار کرد؟

✚ حداکثر تعداد شکار چه میزانی است؟

✚ و ...

فرض می کنیم او مجوز شکار یک اورپال (نوعی آهو) را از شکاربانی منطقه دریافت می کند.

- کار اسپایدر: او اورپالی رعنا را شکار می کند و سپس آنرا با خود به منزل می برد.

- کار ایندکسر: شکاررا تکه تکه کرده، گوشت، استخوان، دل و قلوه، کله پاچه و ... آنرا بسته بندی می کند و بخش های زائدشکار را دور می ریزد.

- کار پایگاه داده: بسته های حاصل را درون فریزر قرارداده، ذخیره می کند.

- کار سیستم رتبه بندی: مهمانان سراغ او می آیند و همسراو بسته به ذائقه مهمانان برای آنها غذا طبخ می کند. ممکن است عده ای کله پاچه، عده ای آبگوشت، عده ای جگر و ... دوست داشته باشند. پخت غذا طبق سلیقه مهمانان کارسختی است. ممکن است همه آنها آبگوشت بخواهند اما آنها مسلماً بامزه ترین آبگوشت را می خواهند!

نکته ها:

✚ شکارچی می توانست برای شکار کبک یا اورپال و یا هر دو به آن منطقه برود، همانطور که موتور جستجوگر می -
تواند از سرور سایت شما انواع فایل (عکس، فایل متنی، فایل اجرایی و ...) درخواست کند.

✚ غذای خوشمزه را می توانید با نتایج جستجوی دقیق و مرتبط مقایسه کنید. اگر شکارچی بهترین شکار را با
خود به منزل ببرد اما غذایی خوشمزه و مطابق سلیقه مهمانان طبخ نکرده، تمام زحمات هدررفته است.

۶. دسته بندی موتورهای جستجو و نحوه کار آنها

موتورهای جستجو به دودسته کلی تقسیم می شوند. موتورهای جستجوی پیمایشی و فهرستهای تکمیل دستی. هر کدام از آنها برای تکمیل فهرست خود از روشهای متفاوتی استفاده می کنند که هر یک را بطور جداگانه مورد بررسی قرار می دهیم:

۱. موتورهای جستجوی پیمایشی

موتورهای جستجوی پیمایشی (Crawler-Based Search Engines) مانند Google لیست خود را بصورت خودکار تشکیل می دهند. آنها وب را پیمایش کرده و سپس کاربران آنچه را که می خواهند از میانشان جستجو می کنند. اگر شما در صفحه وب

خود تغییراتی را اعمال نمایید، موتورهای جستجوی پیمایشی آنها را به خودی خود می یابند و سپس این تغییرات لیست خواهند شد. عنوان، متن و دیگر عناصر صفحه، همگی شامل این لیست خواهند بود.

۲. فهرستهای تکمیل دستی

یک فهرست تکمیل دستی (Human-Powered Directories) مانند یک Open Directory مثل Dmoz وابسته به کاربرانی است که آنرا تکمیل می کنند. شما صفحه مورد نظر را به همراه توضیحی مختصر در فهرست ثبت می کنید یا این کار توسط ویراستارهایی که برای آن فهرست در نظر گرفته شده انجام می شود.

عمل جستجو در این حالت تنها بر روی توضیحات ثبت شده صورت می گیرد و در صورت تغییر روی صفحه وب، روی فهرست تغییری بوجود نخواهد آورد. چیزهایی که برای بهبود یک فهرست بندی در یک موتور جستجو مفید هستند، تاثیری بر بهبود فهرست بندی یک دایرکتوری ندارند. تنها استثناء این است که یک سایت خوب با پایگاه داده ای با محتوای خوب، شانس بیشتری به نسبت یک سایت با پایگاه داده ضعیف دارد.

۳. موتورهای جستجوی ترکیبی با نتایج مختلط

به موتورهایی اطلاق می شود که هر دو حالت را در کنار هم نمایش می دهند. غالباً، یک موتور جستجوی ترکیبی در صورت نمایش نتیجه جستجو از هریک از دسته های فوق، نتایج حاصل از دسته دیگر را هم مورد توجه قرار می دهد. مثلاً موتور جستجوی MSN بیشتر نتایج حاصل از فهرستهای تکمیل دستی را نشان می دهد، اما در کنار آن نیم نگاهی هم به نتایج حاصل از جستجوی پیمایشی دارد.

۴. بررسی یک موتور جستجوی پیمایشی

موتورهای جستجوی پیمایشی شامل سه عنصر اصلی هستند. اولی در اصطلاح عنکبوت (Spider) است که پیمایشگر (Crawler) هم نامیده می شود. پیمایشگر همین که به یک صفحه می رسد، آنرا می خواند و سپس لینکهای آن به صفحات دیگر را دنبال می نماید. این چیز است که برای یک سایت پیمایش شده (Crawled) اتفاق افتاده است. پیمایشگر با یک روال منظم، مثلاً یک یا دو بار در ماه به سایت مراجعه می کند تا تغییرات موجود در آنرا بیابد. هر چیزی که پیمایشگر بیابد به عنصر دوم یک موتور جستجو یعنی فهرست انتقال پیدامی کند. فهرست اغلب به کاتالوگی بزرگ اطلاق می شود که شامل لیستی از آنچه است که پیمایشگر یافته است. مانند کتاب عظیمی که فهرستی را از آنچه که پیمایشگرها از صفحات وب یافته اند، شامل شده است. هرگاه سایتی دچار تغییر شود، این فهرست نیز به روز خواهد شد.

از زمانی که تغییری در صفحه‌ای از سایت ایجاد شده تا هنگامی که آن تغییر در فهرست موتور جستجو ثبت شود، مدت زمانی طول خواهد کشید. پس ممکن است که یک سایت پیمایش شده باشد اما فهرست شده نباشد. تا زمانی که این فهرست بندی برای آن تغییر ثبت نشده باشد، نمی‌توان انتظار داشت که در نتایج جستجو آن تغییر را ببینیم. نرم‌افزار موتور جستجو، سومین عنصر یک موتور جستجو است و به برنامه‌ای اطلاق می‌شود که بصورت هوشمندانه‌ای داده‌های موجود در فهرست را دسته‌بندی کرده و آنها را بر اساس اهمیت طبقه‌بندی می‌کند تا نتیجه جستجو با کلمه‌های درخواست شده هر چه بیشتر منطبق و مربوط باشد.

موتورهای جستجو چه مقداری از کل صفحات وب را برای پیدا کردن پاسخ، جستجو می‌کنند؟

یک محدودیت مهم در موتورهای جستجو یا دایرکتوری‌ها این است که تنها درصد کوچکی از Web توسط آنها برای پیدا کردن صفحات مربوطه به کلمه مورد جستجو بررسی می‌گردد. اگرچه برنامه‌های رهیاب (bot) اطلاعات بیشتری نسبت به دایرکتوری‌ها درباره مورد جستجو جمع‌آوری می‌کنند، با این وجود باز هم مقدار بسیار کوچکی از کل Web تحت جستجوی این برنامه‌ها قرار می‌گیرد. تحقیقات جدید درباره موتورهای جستجو که جدیداً توسط موسسه تحقیقاتی NEC انجام گرفته است بیان می‌کند که حتی بزرگترین موتورهای جستجو کمتر از ۲۰٪ کل وب را مرور می‌کنند. برنامه‌نویسان و طراحان All the Web, Excite و دیگر موتورهای جستجو شدیداً تلاش می‌کنند تا درصد بررسی صفحات وب و همچنین میزان وابستگی موضوع صفحات با مورد جستجو را بالا ببرند.

علت تفاوت حاصل از موتورهای جستجوی گوناگون متفاوت

دلیل تفاوت در نتایج، استفاده کردن موتورهای جستجو از برنامه‌های رهیاب (bot) متفاوت، می‌باشد. هر برنامه Bot در نوع اطلاعاتی که از صفحات وب جمع‌آوری می‌کند و یا به عبارتی «خزیدن» بر روی صفحات وب روش خاصی دارد. بعضی از برنامه‌های رهیاب قبل از این که به آدرس بعدی بروند تمام اطلاعات موجود در صفحه جاری را بررسی می‌کنند. بعضی تنها پیوندهای (Link) موجود در روی صفحه را بررسی می‌کنند. برخی برای بررسی اطلاعات وارد شده جدید، هر روز و یا هر هفته و هر ماه صفحات Web را چک کنند. اطلاعاتی که برنامه رهیاب منعکس می‌کند، از نظر لغات کلیدی (keyword) اطلاعات مربوط به آدرس پایگاه و تازگی صفحات مربوطه متغیر می‌باشد.

مقایسه‌ی کارایی چند موتور جستجو کاربردی

توسعه‌ی اینترنت و میزان اطلاعات روز افزون تولید شده در آن، مسأله بازیابی اطلاعات را هر چه بیشتر مطرح کرده است. برای حل این مسأله، ابزارهای مختلفی پایه عرصه‌ی وجود نهاده‌اند که یکی از کارآمدترین آنها موتورهای جستجو می‌باشد.



مؤسسه‌ها و شرکتهای مختلف، موتورهای جستجوی گوناگونی طراحی نموده‌اند و به‌جرات می‌توان گفت بیش از چهارهزار موتور جستجو تولید شده است که با توجه به تنوع مدارک و اطلاعات، هریک نوع خاصی از اطلاعات را بازیابی می‌کنند. حال برای آشنایی بیشتر با این ابزارهای جستجو، پنج موتور جستجو که کاربرد بیشتری دارند، معرفی و بایکدیگر مقایسه می‌شوند.

این موتورها عبارتند از: Google, Yahoo, Excite, Lycos, Altavista

Altavista: www.altavista.com یا www.altavista.digital.com

این موتور جستجو در سال ۱۹۹۵ توسعه پیدا کرد. نرم‌افزاری است با قابلیت‌های مختلف و جستجوی کلیدواژه‌ای. در قسمت Advanced Query می‌توانید چندین کلمه را با استفاده از عملگرهای Near, Not, Or, And ترکیب کنید. در جستجو می‌توان از کوتاه‌سازی واژه‌ها (Wild Card) استفاده کرد. نتایج جستجو همراه با چندسطری از ابتدای عنوان هر مدرک بازیابی شده آورده می‌شود. همچنین، این موتور جستجو دارای راهنمای کمکی درباره چگونگی استفاده از امکانات آن است.

Lycos: www.lycos.com

یکی از موتورهای به‌شمار می‌رود که قادر است علاوه بر جستجوی تک‌واژه‌ای، واژه‌های ترکیبی را با استفاده از عملگرهای Not, Or, And جستجو کند. اما، قادر به جستجوی جمله یا عبارت نیست. نتایج جستجو با چکیده‌ای کوتاه همراه می‌باشد. علاوه بر جستجوی مدرک، توانایی جستجوی تصاویر و فایل‌های صوتی را نیز داراست.

Excite: www.excite.com

در این نرم‌افزار می‌توان علاوه بر جستجوی کلیدواژه‌ای، جستجوی مفهومی انجام داد. به این صورت که اطلاعاتی را که درباره موضوع مورد نظر شماست، بازیابی می‌کند و به‌طور خودکار مترادفها و واژه‌های مرتبط با موضوع را مورد جستجو قرار می‌دهد. نتایج جستجو همراه با چکیده‌ای از محتوای سایتها است. استفاده از عملگرهای And, Or, Not, و همین‌طور علائمی نظیر (+) و (-) امکان‌پذیر است.

Yahoo: www.yahoo.com

این موتور جستجو برای یافتن اطلاعات علمی چندان مناسب نیست. اما برعکس، بهترین ابزار برای یافتن اطلاعات عمومی در اینترنت است. دارای ساختار سلسله‌مراتب موضوعی است. نتایج جستجو بصورت رده‌های موضوعی طبقه‌بندی شده است. دارای قابلیت جستجوی کلیدواژه‌ای و کوتاه‌سازی واژه‌ها و نیز استفاده از عملگرهای Or و And است.

Google: www.google.com

یکی از کارآمدترین موتورهای جستجو است که نسبت به دیگر موتورها قدمت کمتری دارد. دارای صفحه‌ای ساده و بدون تبلیغات تجاری است. به همین دلیل، بسیار سریع بازیابی می‌شود. دارای امکانات جستجوی ساده و پیشرفته است. همچنین، دارای امکانات راهنمایی (Directory) است که در آن اطلاعات مختلف به صورت طبقه‌بندی شده ارائه شده است.



در جستجو هایی که از چند کلیدواژه استفاده می شود عملگر And به عنوان پیش فرض جستجو محسوب می شود، یعنی در صورت بکاربردن آن، خود موتور آن را در نظر می گیرد. برای جستجوی دقیق تر می توان عبارت مورد نظر را داخل گیومه « » قرار داد تا به طور دقیق، همان عبارت با همان ترتیب کلمات جستجو شود. نتایج جستجو با چکیده ای از محتوای اطلاعاتی سایتها همراه است و در ضمن، در برابر حروف بزرگ حساس نیست. به عبارت دیگر، جستجوی یک واژه با حروف کوچک و بزرگ نتیجه یکسانی دارد. برای مقایسه بهتر موتورهای جستجو، ویژگی های هر یک در جدول زیر آورده شده است:



انواع موتور ها امکانات	Altavista	Lycos	Excite	Yahoo	Google
حوزه جستجو	سایتهای وب سایتهای خبری	سایتهای وب-اخبار مشاغل-آب و هوا- چند رسانه ایها	سایتهای وب- اخبار-تصاویر- فایلهای صوتی (MP3)- آگهی های طبقه بندی شده	نمایه یاهو- گروههای خبری- آدرسهای URL e-mail- فیلد عناوین	سایتهای وب- اخبار مدارک تصویری (pdf) اسناد دولتی
نوع جستجو	کلیدواژه های	کلیدواژه های	کلیدواژه های مفهومی	کلیدواژه های	کلیدواژه های
سطح جستجو	ساده - پیشرفته - جستجوی عبارتی کوتاه سازی واژه با علامت *	ساده - پیشرفته کوتاه سازی انتهای واژه با علامت \$	ساده - پیشرفته	ساده - پیشرفته جستجوی عبارتی کوتاه سازی واژه با علامت *	ساده - پیشرفته - جستجوی عبارتی
استفاده از عملگرهای بولی و علائم دیگر	Or ,And Not ,Near	امکانات کامل عملگرهای بولی	Not, Or, And not,And و علائم (+) و (-)	Or, And	امکانات کامل عملگرهای بولی
چکیده ای از محتوای اطلاعاتی هر یک از سایتهای باز یابی شده	دارد	دارد	دارد	ندارد	دارد
امتیاز داده شده بر اساس گزارش منابع اینترنتی	B+	B+	B+	A	A

جدول ۱: مقایسه موتورهای جستجو

مقایسه برخی از ویژگیهای چند موتور جستجو کاربردی

همانطور که در جدول فوق مشاهده می شود، موتورهای جستجو در بسیاری از موارد شباهتها و در برخی جنبهها تفاوتی با یکدیگر دارند. در زمینه بررسی انواع موتورهای جستجو و مقایسه قابلیت های آنها تحقیقاتی چند صورت گرفته است و در هر یک معایب و محاسن چند موتور برگزیده بیان شده است. اما بطور کلی انتخاب یک موتور جستجو از میان دیگر موتورها بستگی به نوع اطلاعات مورد نیاز فرد دارد.

راهکارهایی برای کوتاه کردن مسیر جستجو

حتی بهترین لغات کلیدی برای جستجوی یک مطلب در محیط اینترنت باعث به دست آمدن تعداد بسیار زیادی جواب شده که امکان بررسی همه آنها برای کاربر تقریباً غیرممکن است. اما چند روش برای کوتاه کردن مسیر جستجو یعنی پیدا کردن نزدیکترین جوابها برای موضوع مورد جستجو وجود دارد. استفاده از عملگرهای منطقی، یک راه حل بسیار معمول و قابل استفاده برای این مشکل است. یعنی با استفاده از لغات AND, OR, NOT در میان لغات کلیدی مطمئن شوید که نتایج به دست آمده حاوی همه، حداقل یکی یا هیچکدام از لغات وارد شده به عنوان موضوع جستجو است. برای مثال یک جستجو برای اتومبیل های روی اینترنت دارای جواب های بسیار زیادی می باشد اما اگر ما بخواهیم به دنبال پایگاه هایی که درباره اتومبیل و منحصرأ BMW هستند بگردیم می توانیم از لغت کلیدی Car AND BMW استفاده کنیم یا اگر پایگاه هایی را که درباره اتومبیل هایی بجز BMW هستند را بخواهیم، لغت کلیدی مورد نظر BMW NOT Car خواهد بود. علامات و لغات دیگری نیز برای گوناگون کردن مسیر وجود داشته که با توجه به نوع برنامه جستجوگر متغیر است. برای اطلاع از این علامات در محیط جستجوگر می توانید به گزینه Help در خود محیط جستجوگر مراجعه کنید. در ذیل با مهارت های جستجو از طریق اینترنت آشنا شوید :

- (۱) حتی المقدور سعی شود کلمات کلیدی از میان اصطلاحات منحصر به فرد و اسامی خاص انتخاب شود.
- (۲) حتی المقدور از آوردن کلمات عمومی که عناوین بسیاری را در زیر مجموعه خود شامل می شوند، جداً خودداری کنید.
- (۳) همیشه اسم شخص یا نام شی یا هر چیز دیگری را که مدنظر دارید به طور کامل وارد کنید.
- (۴) دقت کنید که اگر موتور جستجو میان حروف بزرگ و کوچک تفاوتی می گذارد، این مسأله را در طرح کلمات کلیدی خود مدنظر داشته باشید.
- (۵) در نظر داشته باشید اگر نتیجه جستجو صفر بود به احتمال زیاد می تواند از یک اشتباه تایپی باشد.
- (۶) اگر املا صحیح و کامل کلمه ای را نمی دانید از کارکتر جانشین که اغلب * و یا ؟ است استفاده کنید.

۷) اگر یک کلمه کلیدی برای طرح دقیق و تمام و کمال یک مورد جستجو کفایت نمی کند، از تکنیک های جستجوی- عبارتی، استفاده از اپراتورهای جبرولین (AND, OR, NOT) استفاده کنید. جستجوی عبارتی یکی از مهم ترین و قدرتمندترین امکانات جستجو در اغلب موتورهای جستجو می باشد و می توان یک عبارت یا جمله مشخص را به همان ترتیبی که کلمات وارد شده اند مورد جستجو قرارداد. برای این روش جستجو عبارت مورد نظر را داخل گیومه « » بگذارید.

۸) استفاده از عملگر AND: به مفهوم « و » برای محدود کردن دامنه جستجو از طریق ترکیب کلیدواژه های مختلف به کار می رود و برای ترکیب کلیدهای جستجو زمانی که برای شما مهم است که دو یا چند کلمه کلیدی حتماً وجود داشته باشد و علامت آن در پایگاه های مختلف به صورت استفاده از عبارت AND، استفاده از +، انتخاب عبارت ALL THE WORD از منو، انتخاب عبارت (MATCH ON ALL WORDS AND) به وسیله کلیک کردن روی دکمه های رادیویی است.

۹) استفاده از عملگر OR: اپراتور OR به مفهوم « یا » و برخلاف عملگر AND باعث گسترش دامنه جستجو و بازیابی اطلاعات بیشتر شده برای ترکیب کلیدواژه های جستجو زمانی که انتظار دارید فقط یک، دو یا چند کلمه کلیدی حضور داشته باشند و علامت آن استفاده از عبارت OR، چگونگی اجرای ساده و معمولی آن، انتخاب عبارت ANY OF THE WORDS از منو، انتخاب عبارت (MATCH ON ANY WORDS OR) با کلیک بر روی دکمه های رادیویی می- باشد. یکی از کاربردهای مهم این عملگر پوشش مفاهیم یا اصطلاحات مترادف، مرتبط یا با املاهای متفاوت است.

۱۰) استفاده از عملگر NOT: اپراتور NOT به مفهوم « نه » و یا به جز که در این صورت همه جواب های بازگشتی که حاوی عبارت یا کلمه کلیدی هستند حذف خواهند گردید و برای اجرای آن فقط کافیست که NOT را قبل از عبارت یا کلمه کلیدی مورد نظرتان با یک فاصله بیاورید.

۱۱) استفاده از کوتاه سازی کلیدواژه ها: این تکنیک به ما امکان می دهد که با وارد کردن بخشی از یک کلیدواژه بتوانیم مشتقات مختلف آن را نیز در فرآیند جستجو بازیابی کنیم. اکثر موتورهای جستجو این تکنیک را با استفاده از علامت ستاره (*) ارائه می دهند. یکی از مشکلات استفاده از این تکنیک این است که باعث بازیابی اطلاعات غیر مرتبط و ناخواسته زیادی می شود.

۱۲) استفاده از عملگر نزدیک یابی: در بسیاری از موارد استفاده از عملگر AND باعث بازیابی اطلاعاتی می شود که برای ما مفید نیست. به این دلیل که این عملگر کلیدواژه ها را در هر کجای متن که باشند بازیابی می کند. در این موارد استفاده از تکنیک نزدیک یابی می تواند ریزش کاذب اطلاعات و یا بازیابی اطلاعات غیر مرتبط جلوگیری نماید. همه موتورهای جستجو قابلیت استفاده از این تکنیک را ندارند ولی به عنوان مثال در موتور جستجوی آلتاویستا می توان با استفاده از عملگر NEAR از این تکنیک استفاده نمود.

۱۳) جستجوی ترکیبی با استفاده از پیرانتز: این تکنیک یکی از مهم ترین تکنیک های جستجو می باشد که به وسیله آن می توان تا حدود زیادی از بازایی موارد غیر مرتبط در محیط وب جلوگیری کرد. در این روش می توان از همه عملگرهای جستجو که در بالا گفته شده یک جا استفاده کرد و آن ها را با هم دیگر ترکیب نمود.

۱۴) جستجوی کلیدواژه در عنوان صفحات وب: این تکنیک با این پیش فرض که عنوان یک صفحه وب تا حدود زیادی نمایان گر محتوای اطلاعات موجود در آن است به جستجوی واژه های کلیدی در عنوان سایت ها می پردازد. علامت آن در موتورهای جستجو متفاوت است ولی اغلب موتورهای جستجو از طریق فهرست انتخابی و یا گزینه های دیگر این امکان را فراهم می آورند.

۱۵) جستجوی حوزه سایت ها: با توجه به این که به صورت قراردادی هر کشوری حوزه خاصی در محیط وب دارد، قابلیت- جستجوی حوزه سایت ها به ما این امکان را می دهد که فرایند جستجو را به حوزه خاصی نظیر سایت های وب ایران (IR) و یا سایت های وب سازمان های غیرانتفاعی (ORG) محدود کنیم. دستورات استفاده از این تکنیک در موتورهای جستجو مختلف می باشد.

۱۶) محدود کردن جستجو به زبان های مختلف باعث می شود نتایج جستجو به زبان های دیگر آورده نشود و انتخاب مطلب مورد نظر آسان تر است.

۱۷) محدود کردن جستجو به تاریخ انتشار منابع در وب: تاریخ انتشار یا به اصطلاح روزآمدی مطلب به خصوص در منابع- علمی اصل مهمی است و این گونه محدودیت باعث می شود بنابه نیاز کاربر جدیدترین و یا قدیمی ترین منبع بازایی بشود.

۱۸) جستجوی رسانه های مختلف؛ موسیقی، عکس، ویدئو؛ زمانی که فقط نوع خاصی از رسانه مورد نیاز است به عنوان مثال زمانی که به عکس یک شخصیت نیاز داریم، جستجو در میان عکس ها باعث می شود نتیجه جستجو شامل اطلاعات دیگری در مورد آن شخصیت نباشد.

۱۹) جستجوی صفحات با فرمت های مختلف؛ PDF, WORD, MP3, MPEG: زمانی که فرمت خاصی مورد نظر است می توان از این تکنیک استفاده کرد. به عنوان مثال اگر مایل باشیم منبع بازایی شده در فرمت PDF باشد، این تکنیک می- تواند مفید باشد.

۲۰) آگاهی از پیش فرض های جستجو در موتور جستجو: با توجه به این که هر موتور جستجو برای ترکیب واژه ها یک پیش- فرض دارد و اگر از هیچ گونه عملگری استفاده نشود، کلیدواژه ها را به صورت پیش فرض با یکی از عملگرهای جبر بولی ترکیب می کند؛ آگاهی از این پیش فرض موتورهای جستجوی مختلف مهارت ما را در جستجو بالا می برد.

۲۱) وب نامرئی: وب نامرئی به دودلیل کمی و کیفی اهمیت دارد. کمی از این نظر که موتورهای جستجو فقط قادر هستند حدود ۱۶ درصد از اطلاعات موجود در اینترنت را بازیابی کنند و اندازه وب نامرئی تقریباً ۵۰۰ برابر وب مرئی است و کیفی از این نظر که منابع اطلاعاتی موجود در وب عمیقاً معمولاً ارزشمند و مفید هستند و در بسیاری از موارد پاسخگوی نیاز کاربران می باشند. آشنایی با ابزارهایی که برای شناسایی منابع وب نامرئی به وجود آمده اند و کاربران را به سایت های مناسب راهنمایی می کنند، باعث دسترسی به این بخش عظیم از اطلاعات مفید و ارزشمند می شود. مثل سایت INVISIBLEWEB که فهرستی از منابع نامرئی را، و سایت COMPLETEPLASSET که فهرستی از تقریباً ۴۰۰۰۰ پایگاه اطلاعاتی وب نامرئی را ارائه می دهد.

مهم ترین ترندهای گوگل

۱- مستثنی کردن کلمات

بامستثنی کردن کلمات، علاوه بر کاهش حجم نتایج جستجو، می توان به اطلاعات مورد نظر نزدیک تر شد. با گذاشتن علامت «-» قبل از کلمه دلخواه، گوگل آن کلمه را استثناً قلمداد کرده و از نمایش صفحاتی که حاوی چنین کلمه ای باشند، پرهیز می کند. مثلاً نتیجه عبارت، همه صفحاتی که مربوط به ایران هستند و از تهران نامی نبرده اند را شامل می شود.

۲- WILDCARD ها

زمانی که به طور کامل از کلمات موضوع مورد جستجو آگاهی ندارید، بهتر است از کاراکتر کمکی ستاره * استفاده کنید. کاراکتر * معنی هرچیز می دهد. مثلاً عبارت می تواند نتایجی نظیر، و یا نتایجی از این دست را حاصل شود.

۳- عناوین صفحات

معمولاً همه صفحات وب، عنوان دارند. گاهی اوقات استفاده از این عناوین، به شما در پیدا کردن مطالب مورد نیازتان کمک شایانی می کند. مثلاً عبارت زیر، گوگل را بر آن می دارد تا فقط عناوین صفحات را برای یافتن کلمه sanobar جستجو کند:

sanobar :INTITLE

۴- شرکت در گروه های خبری

اینترنت فراتر از ایمیل و صفحات وب است. به عنوان مثال گروه های خبری (NEWS GROUP) یکی دیگر از خدماتی است که از طریق اینترنت ارائه می شود. گروه های خبری گوگل، گروه های معروف و معتبری هستند که شرکت در آن ها، به-

نوبه خود، کار لذت بخشی است. آدرس گروه های خبری گوگل از این قرار است: GROUPS.GOOGLE.COM



۵- جستجوی خودکار

جستجو در اینترنت، قصه سوزن و انبارکاه است. گاهی اوقات ممکن است علیرغم صرف وقت زیاد، بازهم به نتیجه دلخواه نرسید. بنابراین هر ابزاری که در این پروسه، از اتلاف وقت جلوگیری کند، بسیار باارزش است. یکی از این ابزارها، «پیغام‌دهنده» گوگل نام دارد. این ابزار با دریافت آدرس ایمیل و عبارت مورد جستجو، به‌طور خودکار و روزانه، جستجویی را براساس عبارت‌دریافتی ترتیب‌داده و هر زمان که اطلاعات جدیدی را یافت‌نماید، آدرس آن را برایتان ایمیل می‌کند. آدرس پیغام‌دهنده گوگل به این ترتیب است:

[/HTTP://WWW.GOOGLEALERT.COM](http://www.googlealert.com)

۶- جستجوگر اخبار

این ابزار نیز همانند پیغام‌دهنده گوگل، مشخصات اخبار مورد نظر را به‌همراه آدرس ایمیل دریافت کرده و روزانه (بنا به درخواست کاربر، قابل تنظیم است). براساس عبارت‌دریافتی، جستجویی را در سایت‌های خبری ترتیب می‌دهد و در صورت یافتن اطلاعات مناسب، آن را برایتان ایمیل می‌کند.

آدرس این ابزار <http://www.google.com/newsalerts> است.

۷- ماشین حساب گوگل

گوگل برای حل معادلات ریاضی هم، ابزاری دارد که بسیار سریع و بادقت عمل می‌کند. این ابزار علاوه بر حل معادلات ریاضی، قابلیت تبدیل انواع واحدها به یکدیگر را نیز دارد. برای آشنایی بیشتر با این ابزار کارآمد به آدرس زیر مراجعه کنید:

[HTTP://WWW.GOOGLE.COM/HELP/FEATURES.HTM#CALCULATOR](http://www.google.com/help/features.htm#calculator)

۸- معنی کلمات

اگر معنی یا بهتر بگوییم، تعریف کلمه‌ای را نمی‌دانید، در کادر جستجو کلمه مورد نظر: DEFINE را نوشته و کلید ENTER را بزنید. در جواب این جستجو، علاوه بر تعریف کلمه، لینکی نیز نمایش داده می‌شود که با کلیک روی آن، به جزئیات بیشتری می‌رسید.

۹- ترجمه کنید

در حال حاضر، ماشین ترجمه گوگل می‌تواند انگلیسی را به ۵ زبان زنده دنیا ترجمه کند (امکان ترجمه بالعکس نیز وجود دارد). برای استفاده از این مترجم ماشینی به آدرس زیر مراجعه نمایید:

[HTTP://WWW.GOOGLE.COM/LANGUAGE_TOOLS?HL=EN](http://www.google.com/language_tools?hl=en)



۱۰- تصویر یاب

این ابزار بادریافت کلمه یا عبارتی، تصاویر موجود روی اینترنت را جستجو کرده و در نهایت آنهایی را که با عبارت دریافتی منطبق هستند، فهرست می کند. برای استفاده از این ابزار به آدرس <http://WWW.GOOGLE.COM/IMGHP> مراجعه نمایید.

۱۱- پرسش و پاسخ

تحقیق و جمع آوری مطالب از اینترنت، کار وقت گیری است. اگر برای انجام این کار وقت کافی در اختیار ندارید، سری به آدرس [HTTP://ANSWERS.GOOGLE.COM/ANSWERS](http://ANSWERS.GOOGLE.COM/ANSWERS) بزنید و بامطرح کردن سوالاتتان و بررسی جوابها، سریع تر به هدف تان برسید.

۱۲- نمایش پنجره تنظیمات

گوگل به شما این امکان را می دهد تا نحوه نمایش نتایج را مطابق میل تان تنظیم کنید. به عنوان مثال بامراجعه به آدرس <http://WWW.GOOGLE.COM/PREFERENCES> می توانید تعداد نمایش نتایج در هر صفحه را تعیین نمایید. علاوه بر این در قسمت تنظیمات امکانات دیگری نظیر SAFE SEARCHING نیز وجود دارد. با فعال کردن این گزینه، گوگل از نمایش سایت هایی که به هردلیلی غیرمجاز هستند، پرهیز می کند.

۱۳- آمار استفاده از گوگل

در سایت <http://WWW.GOOGLE.COM/PRESS/ZEITGEIST.HTML> آمار جهانی استفاده از گوگل موجود است. در این سایت جزییاتی از چگونگی استفاده ملل مختلف از گوگل و الگوهای جستجوی آن ها وجود دارد.

۱۴- جستجو در دسته بندی ها

علاوه بر جستجوی کلی وب، گوگل امکان جستجو در دسته بندی ها (CATEGORIES) را نیز می دهد. به عنوان مثال برای یافتن اطلاعاتی درباره هنر، تجارت، ورزش و ... می توانید ضمن مراجعه به نشانی: <http://WWW.GOOGLE.COM/DIRHP> جستجویی را در دسته بندی های گوگل ترتیب بدهید و سریع تر به هدف تان برسید.

۱۵- برای بلاگرها

وبلاگ داشته باشیم یا نه؟ ایده داشتن یک وبلاگ شخصی، بسیار جالب است و این که با در اختیار داشتن یک وبلاگ- شخصی حرف دلمان را به گوش جهانیان برسانیم، فکر بسیار خوبی است. برای ساخت یک وبلاگ شخصی، به آدرس

<http://WWW.BLOGGER.COM/START> مراجعه نمایید. این سایت متعلق به گوگل است.

۱۶- جستجوی داخل فایل ها

با کمک گوگل می توان داخل فایل های مشخصی را برای یافتن کلمه ای دلخواه جستجو کرد. مثلاً عبارت IRAN FILETYPE:DOC گوگل را بر آن می دارد تا تمامی فایل های DOC (مربوط به برنامه MICROSOFT WORD) را به منظور یافتن واژه IRAN جستجو کرده و آن هایی که دارای چنین کلمه ای هستند را فهرست کند.

۱۷- کلمات عمومی

برخی کلمات عمومی نظیر AT, TO, THE و غیره، زمانی که در عبارت مورد جستجو ذکر می شوند، توسط گوگل نادیده انگاشته می شوند و تاثیری در نتیجه ندارند. حال اگر می خواهید چنین کلماتی در نتیجه جستجو موثر باشند، کافی است یک علامت « + » در مقابل آن ها بگذارید.

۱۸- موتورهای جستجوی تخصصی

باین ترند می توانید ضمن معرفی منابع مورد نظران به گوگل، از آن بخواهید که نتایج جستجو را فقط از همان منابع بیاورد. مثلاً برای جستجو در سایت مایکروسافت به GOOGLE.COM را انتخاب کنید. گوگل تمامی صفحاتی که در سه ماه گذشته ایجاد یا بروز شده اند را جستجو و آن هایی که با موضوع مورد نظران منطبق هستند را نمایش می دهد.

۷. بررسی کمی و کیفی موتورهای جستجو

گوگل با ادعای ایندکس کردن ۸ میلیارد صفحه سهم خود را در مبارزه بین اندازه موتورهای جستجوگر افزایش داده است. مایکروسافت هم که قصد دارد عنوان بزرگترین موتور جستجوگر جهان را کسب کند امروز ۵ میلیارد صفحه را ایندکس کرده است. این مسئله موجب می شود تا از ۴.۲ میلیارد صفحه ای به گزارش خود گوگل در سال گذشته ایندکس شده است پیشی بگیرد. ما پیشتر از این نیز در جریان جنگ اندازه قرار داشتیم.

این نزاع وقتی بوجود می آید که یکی از موتورهای جستجو در تلاش برای بدست آوردن مشخصه مثبتی باشد که آنرا نسبت به سایرین در جایگاه بالاتری قرار می دهد. البته باید توجه داشت که بزرگی اندازه دلیلی برای اثبات برتری نیست. یک موتور جستجو به همراه شمارزادی از صفحات، اگر به موقع بروزرسانی نشود و یا ارتباطات صحیح وجود نداشته باشد ممکن است حقیقتاً ضعیف تر از موتور جستجویی باشد که صفحات کمتری را در بردارد. بطور خلاصه در مورد میزان کمیت و

کیفیت موتورهای جستجو دو معیار میزان صفحات ایندکس شده و میزان عمق صفحات بسیار مهم می باشند. در ادامه این دو معیار را به تفصیل بررسی می کنیم.

۱. اندازه گزارش شده

این عدد (Reported Size) درحقیقت آن چیزی است که موتورهای جستجو ادعا می کنند. درموردگوگل این عدد برخی اوقات شامل صفحات ایندکس شده نسبی "partially-indexed" و یا به معنای روشن تر صفحات فقط لینک (link-only) است. اینها صفحاتی بوده اند که گوگل منحصراً از طریق لینک می شناسد و خود این صفحات درواقع ایندکس نشده اند. نوعاً، موتورهای جستجو قادر به شناسایی صفحات تکراری، صفحات اسپم و نظیر آن نیستند. البته ما اینجا به این مسئله نمی پردازیم و فرض می کنیم چنین قابلیتی دارند.

اما یاهو سعی می کند از بازی اندازه صفحات خودش را بیرون بکشد. اوایل امسال وقتی تکنولوژی جدید جستجوی خود را پیاده سازی می کرد، از اعلام میزان صفحات ایندکس شده خودداری کرد و به گفتن این جمله که «موتور جستجوی ما قابل مقایسه با دیگران است» اکتفا کرد. سخنگوی یاهو در این باره می گوید: «همانند گذشته یاهو سایز صفحات ایندکس شده خود را به دلایل رقابتی ذکر نمی کند. ما اعتقاد داریم ایندکس ما کاملاً قابل رقابت با دیگران است».

کیفیت جستجو شامل فاکتورهای متفاوتی نظیر بروزسانی، میزان وابستگی و ... است. ما تلاش می کنیم تا نتایج با کیفیت بالاتری برای مصرف کنندگانمان برسیم تا به آنها اطمینان دهیم که قادر خواهند بود بهترین نتیجه مورد نظرشان را از طریق موتور جستجوی ما بدست بیاورند. این رویه یاهو هم خوشایند است هم ناخوشایند. خوشایند از آن جهت که نگفتن این عدد مانع از وارد شدنشان به بحث بیهوده اندازه صفحات ایندکس شده می شود. ناخوشایند از آن که نگفتن این عدد موجب می شود تا آمار دقیقی از عملکرد آنها نداشته باشیم. به نظر می رسد همه موتورهای جستجوگر باید آنرا بدون ریاکاری ذکر کنند. با این حساب می توان میزان صفحات ایندکس شده یاهو را با توجه به ادعای قابل رقابت بودن موتور جستجوی آنها ۴.۲ میلیارد صفحه فرض کرد. این رقم معادل میزان صفحات چند ماه پیش گوگل است.

۲. عمق صفحات

بررسی عمق صفحات (Page Depth Amount) ایندکس شده بسیار جالب تر خواهد بود. فرض کنید که شما میلیون صفحه را می خواهید ایندکس کنید. آیا حقیقتاً کل متن صفحات را ایندکس می کنید یا بخشی از آنرا؟ با استفاده از این معیار می توان فهمید که عملکرد یک موتور جستجو چگونه است. درمورد گوگل تقریباً می توان گفت که بعضی از صفحات را تنها بصورت نسبی و بخشی از آنرا ایندکس می کند. در گذشته اگر اندازه صفحه ای بیشتر از K101 بود تنها K101 اول متن

توسط گوگل ایندکس می شد و بقیه متن کنار گذاشته میشد. برطبق فرضیات دقیق، گوگل هنوز هم به همین شکل عمل می کند.

میزان عمق صفحات ایندکس شده MSN را می توان از گفته هایی که در آگوست سال گذشته به هنگام جلسه ملاقات کرولرها (Crawlers) در همایش موتورهای جستجو درس خوزه بیان شد فهمید. البته در مورد نسخه جدید جستجوگر MSN شاید شرایط فرق کرده باشد. میزان این معیار در یاهو نیز در همان جلسه ای که ذکر شد بیان شد. اما مدیران Jeeves Ask از گفتن میزان عمق صفحاتشان در جلسه مذکور خودداری کردند و به جمله « ما در حد دیگران هستیم » قناعت کردند. لذا می توان عمق صفحات آنها را چیزی شبیه به گوگل دانست. بررسی پیدا کردن میزان این معیار در موتورهای جستجو بسیار آسان است. برای این کار کافی است یک صفحه حجیم که طول زیادی دارد را پیدا کنید و به جستجوی مطلبی که در انتهای آن - صفحه قرار دارد بپردازید. نتایج بدست آمده را مقایسه کنید تا ببینید کدام جستجوگر عمق پایین تر را بهتر سرچ می کند. در بررسی نتیجه یک تحقیق مشخص شد که یاهو حقیقتاً قادر است تا عمق ۸۰۰K صفحات را نیز جستجو کند. در پایان باید گفت هر چند بیشتر بودن معیار میزان صفحات ایندکس شده یک موتور جستجوگر مفید است، اما هرگز به معنای مرتبط بودن و درست بودن صفحات ایندکس شده نیست. هر چند گوگل تقریباً دو برابر یاهو صفحه ایندکس کرده - است، اما به این معنی نیست که دو برابر یاهو بهتر و کارآمدتر است.

چگونه موتورهای جستجو صفحات وب را رتبه بندی می کنند؟

وقتی شما از موتورهای جستجوی پیمایشی چیزی را برای جستجو درخواست می نمایید، تقریباً بلافاصله این جستجو از میان میلیونها صفحه صورت گرفته و مرتب می شود، بطوریکه مربوطترین آنها نسبت به موضوع مورد درخواست شما رتبه - بالاتری را احراز نماید. البته باید در نظر داشته باشید که موتورهای جستجو همواره نتایج درستی را به شما ارائه نخواهند داد و مسلماً صفحات نامربوطی را هم در نتیجه جستجو دریافت می کنید و گاهی اوقات مجبور هستید که جستجوی دقیق تری را برای آنچه که می خواهید انجام دهید اما موتورهای جستجو کارحیرت انگیز دیگری نیز انجام می دهند.

فرض کنید که شما به یک کتابدار مراجعه می کنید و از وی درباره «سفر» کتابی می خواهید. او برای اینکه جواب درستی به شما بدهد و کتاب مفیدی را به شما ارائه نماید با پرسیدن سؤالاتی از شما و با استفاده از تجارب خود کتاب مورد نظرتان را به شما تحویل خواهد داد. موتورهای جستجو همچنین توانایی ندارند اما به نوعی آنها را شبیه سازی می کنند. پس موتورهای جستجوی پیمایشی چگونه به جواب مورد نظرتان از میان میلیونها صفحه وب می رسند؟ آنها یک مجموعه از

قوانین را دارند که الگوریتم نامیده می شود. الگوریتم های مورد نظر برای هر موتور جستجویی خاص و تقریباً سری هستند، اما به هر حال از قوانین زیر پیروی می کنند:

✚ مکان و تکرار

یکی از قوانین اصلی در الگوریتم های رتبه بندی موقعیت و تعداد تکرار کلماتی است که در صفحه مورد استفاده قرار گرفته اند که بطور خلاصه روش مکان - تکرار (Location/Frequency Methode) نامیده می شود. کتابدار مذکور را به خاطر می آورید؟ لازم است که او کتابهای در رابطه با کلمه «سفر» را طبق درخواست شما بیابد. او در وهله اول احساس می کند که شما به دنبال کتابهایی هستید که در نامشان کلمه «سفر» را شامل شوند. موتورهای جستجو هم دقیقاً همان کار را انجام می دهند. آنها هم صفحاتی را برایتان لیست می کنند که در برچسب Title موجود در کد HTML حاوی کلمه «سفر» باشند.

موتورهای جستجو همچنین به دنبال کلمه مورد نظر در بالای صفحات و یا در ابتدای پاراگرافها هستند. آنها فرض می کنند که صفحاتی که حاوی آن کلمه در بالای خود و یا در ابتدای پاراگرافها و عناوین باشند به نتیجه مورد نظر شما مربوط تر هستند. تکرار یا Frequency عامل بزرگ و مهم دیگری است که موتورهای جستجو از طریق آن صفحات مربوط را شناسایی می نمایند. موتورهای جستجو صفحات را تجزیه کرده و با توجه به تکرار کلمه ای در صفحه متوجه می شوند که آن کلمه نسبت به دیگر کلمات اهمیت بیشتری در آن صفحه دارد و آن صفحه را در درجه بالاتری نسبت به صفحات دیگر قرار می دهند.

✚ عوامل خارج از صفحه

موتورهای جستجوی پیمایشی اکنون تجربه فراوانی در رابطه با وب مسترهای دارند که صفحات خود را برای کسب رتبه بهتر مرتباً بازنویسی می کنند. بعضی از وب مسترهای خبره حتی ممکن است به سمت روشهایی مانند مهندسی معکوس برای کشف چگونگی روشهای مکان - تکرار بروند. به همین دلیل، تمامی موتورهای جستجوی معروف از روشهای امتیاز بندی «خارج از صفحه» استفاده می کنند. عوامل خارج از صفحه، عواملی هستند که از تیررس وب مسترها خارجند و آنها نمی توانند در آن دخالت کنند و مسأله مهم در آن تحلیل ارتباطات و لینکهاست. بوسیله تجزیه صفحات، موتورهای جستجو لینکها را بررسی کرده و از محبوبیت آنها می فهمند که آن صفحات مهم بوده و شایسته ترفیع رتبه هستند. علاوه تکنیکهای پیشرفته به گونه ای است که از ایجاد لینکهای مصنوعی توسط وب مسترها برای فریب موتورهای جستجو جلوگیری می نماید. علاوه بر آن موتورهای جستجو بررسی می کنند که کدام صفحه توسط یک کاربر که کلمه ای را جستجو کرده انتخاب می شود و سپس با توجه به تعداد انتخابها، رتبه صفحه مورد نظر را تعیین کرده و مقام آنرا در نتیجه جستجو جابجا می نمایند.

دستور آشنایی

خب آشنایی چه ربطی به موتور جستجو دارد؟ رابطه در اینجاست. همانطور که آشنایان حرفه‌ای، دستورات آشنایی خود را در لفافه نگه می‌دارند و مقدار و چگونگی ادویدهای افزودنی به غذای خود را افشانی نمی‌کنند، چگونگی کارکرد دقیق موتورهای جستجو درباره روشهایی از قبیل مکان - تکرار فاش نمی‌شود و هر موتور جستجویی روش خود را دنبال می‌کند. به همین دلیل است که وقتی شما کلمات واحدی را در موتورهای متفاوت جستجو می‌کنید، به نتایج متفاوتی می‌رسد. برخی موتورهای جستجو نسبت به برخی دیگر صفحات بیشتری را فهرست کرده‌اند. نتیجه این خواهد شد که هیچ موتور جستجویی نتیجه جستجوی مشترکی با موتور دیگر نخواهد داشت و شما نتایج متفاوتی را از آنها دریافت می‌کنید. موتورهای جستجو همچنین ممکن است که برخی از صفحات را از فهرست خود حذف کنند. البته به شرطی که آن صفحات با Spam شدن سعی - در گول زدن موتورهای جستجوگر داشته باشند. Spamming روشی است که برخی از صفحات برای احراز رتبه بالاتر در موتورهای جستجو در پیش می‌گیرند و آن به این صورت است که با تکرار بیش از حد کلمات بطور عمدی سعی در برهم زدن تعادل و در نتیجه فریب موتورهای جستجو را دارند. آنها سعی دارند که با افزایش عامل تکرار، در رتبه بالاتری قرار بگیرند. موتورهای جستجو راههای متنوعی را برای جلوگیری از Spamming دارند و در این راه از گزارشهای کاربران خود نیز بهره می‌برند.

۸. روش کار موتورهای جستجو در رتبه بندی سایتها

وقتی کاربری عبارتی را جستجو می‌کند، موتور جستجو لیستی از سایتها را ارائه می‌دهد که مرتبط با کلمات نوشته شده است. این سایتها بر حسب میزان ارتباط خود با موضوع جستجو شده به صورت نزولی لیست می‌شوند یعنی سایتی که بیشترین ارتباط و بالاترین رتبه را دارد، اولین است.

هر موتور جستجو الگوریتم خاص و محرمانه خودش را برای رتبه بندی (Ranking) دارد و با توجه به اهمیت و اولویت فاکتورهای رتبه بندی ممکن است نتایج جستجوی متفاوتی بوجود آید. ممکن است سایت شما در یک موتور جستجو رتبه ۲ و در موتور دیگر رتبه ۱۰ داشته باشد.

دیگر اینکه رتبه سایت شما کاملاً به کلیدواژه‌هایی که انتخاب می‌کنید بستگی خواهد داشت. برای نمونه ممکن است سایت شما هنگام جستجوی یک عبارت رتبه ۱۲ داشته باشد و در جستجوی دیگری با یک عبارت دیگر، رتبه ۱۲۶۴ را بدست آورد. پس عبارتهایی که برای بهبود جایگاه خود انتخاب می‌کنید مهم است.

بهینه‌سازی رتبه وبسایت در موتورهای جستجو

بهینه‌سازی رتبه وبسایت در موتورهای جستجو روشی است برای بالا بردن تعداد و کیفیت بازدیدکنندگان یک سایت به کمک موتورهای جستجو.

موتورهای جستجو بهترین روش برای یافتن محصولات و اطلاعات مورد نیاز توسط کاربران و متقاضیان از یک سو و از سوی دیگر بهترین راه برای تبلیغات و جذب مشتریان توسط تولیدکنندگان و ارائه‌کنندگان خدمات است.

آمار نشان می‌دهد تنها ۷٪ کاربران از صفحه‌های سوم به بعد نتایج را مرور می‌کنند. به عبارتی، بیشتر آنان اگر مطلب مورد نظر خود را در صفحات نخست پیدا نکنند از موتور جستجوی دیگری کمک می‌گیرند.

سه ابزار اصلی برای بدست آوردن رتبه بالا در موتورهای جستجو و در نتیجه جایگیری در صفحات نخست نتایج جستجو عبارتند از:

۱. بهینه‌سازی موتورهای جستجو

۲. معرفی به موتورهای جستجو و فهرست‌ها

۳. تبلیغات در موتورهای جستجو

بهینه‌سازی موتورهای جستجو

هدف از بهینه‌سازی صفحات وب، جایگیری در صفحات نخست نتایج جستجو است. قرار گرفتن در رده‌های بالای نتایج جستجو می‌تواند صدها و شاید هزاران بازدیدکننده را روزانه به سایت شما راهنمایی کند. این دسته از بینندگان که در پی جستجوهای خود به سایت شما وارد شده‌اند، کاملاً هدفمند به دنبال موضوع مطرح شده در سایت شما بوده‌اند.

فرآیند بهینه‌سازی رتبه وبسایت در موتورهای جستجو شامل مراحل زیر است:

۱. انتخاب کلیدواژه‌ها

کلیدواژه‌ها همان کلماتی هستند که وقتی مورد جستجو قرار می‌گرفتند سایت یا سند ما در صفحات نخست نتایج جستجو (حداقل ۳ صفحه اول) ظاهر شود. در انتخاب کلمات کلیدی باید به چند نکته توجه شود:

✚ ارتباط معنایی با محتویات سایت

کلیدواژه‌ها باید با موضوع سایت مرتبط باشند. اگر این چنین نباشد رتبه سایت بسیار پایین خواهد آمد و حتی ممکن است به عنوان تقلب شناخته شود.

✚ میزان محبوبیت

محبوبیت کلیدواژه به معنای میزان تقاضای آن از طرف کاربران است. هرچه این محبوبیت بیشتر باشد، بدون شک بعد از قرارگیری در رتبه بندی موتور جستجو، ترافیک بیشتری بوجود می آید. دامنه محبوبیت هم از اهمیت بالایی برخوردار است. مثلاً یک واژه آلمانی بین یک عده خاص (آلمانی زبان ها) دارای محبوبیت است.

✚ در نظر گرفتن رقیبان

اگر کلیدواژه شما از محبوبیت بالایی برخوردار است اما رقیبان بسیاری با شما رقابت می کنند (بویژه که از سایت های مهم باشند) بهتر است از آن واژه صرف نظر کنید و یا تا حد امکان تخصصی تر آن را دنبال کنید، یعنی با ترکیب واژه بادیگر واژه ها یک عبارت چند کلمه ای بسازید. کیفیت رقبا شما نیز خیلی مهم است. یعنی آن ها چقدر «مهم» هستند. مثلاً رقابت با سایت ebay خیلی سخت تر و دشوارتر از رقابت با یک وبلاگ محلی است.

۲. بکارگیری کلیدواژه ها در URL

نام دامنه سایت خود را متناسب با محتوای سایتتان انتخاب کنید و کلیدواژه ها را در نام پوشه و نام فایل بکاربرید.

۳. بکارگیری کلیدواژه ها در صفحه

✚ عنوان صفحه (Title)

عنوان صفحه شامل ۵ تا ۱۰ کلمه است و همان متنی است که بالای گردشگر اینترنت ظاهر می شود و موقع ذخیره کردن صفحه و اضافه کردن آن به صفحات محبوب به عنوان اسم پیش فرض ظاهر می شود.

✚ تگ توصیف (Meta Description)

برخی از موتورهای جستجو این متن را که شامل ۲۵ تا ۳۰ کلمه در توصیف صفحه است بعد از عنوان سایت در نتایج جستجوی خود نشان می دهند. از کلیدواژه ها در این متن استفاده کنید.

✚ تگ کلیدواژه ها (Meta Keywords)

این تگ، جزء تگ های بسیار مهم سایت است، زیرا موتورهای جستجو برای این بخش ارزش زیادی قائلند. ابتدا کلیدواژه اصلی، بعد کلیدواژه دوم و بعد چند کلیدواژه مشابه را قرار دهید.

✚ تگ سرصفحه (H1)

کلیدواژه خود را در سرصفحه قرار دهید.

✚ تگ تو پر (B)

بکارگیری کلیدواژه در متن در حالت توپر باعث نتیجه بهتر می شود.

✚ تگ مشخصات عکس (Alt)

وقتی در سایتی ماوس را بر روی عکسی نگه می‌دارید، برای چندثانیه متنی ظاهر می‌شود که شامل توضیحی در مورد عکس است. موتورهای جستجو حتی این توضیحات به‌ظاهر کوچک را در نظر می‌گیرند. به همین دلیل از کلیدواژه‌ها در این متن استفاده کنید.

نام فایل‌های عکس

نام فایل‌های عکس موجود در سایتتان را از میان کلیدواژه‌ها انتخاب کنید و نام را فارسی انتخاب نکنید.

متن صفحه (Content)

ارتباط کامل متن با موضوع و نگارش ساده و صحیح آن مهم است. کلیدواژه‌ها را نیز طوری در متن صفحه بکاربرید که ساختار متن حفظ شود. پارامترهای زیر را در بکارگیری کلیدواژه‌ها در نظر بگیرید:

جایگاه کلیدواژه (Prominence)

منظور میزان نزدیکی کلیدواژه به ابتدای صفحه است. اولویت به ترتیب در مکان‌های زیر است: عنوان متن، عنوان فصل‌ها، اولین کلمه پاراگراف، خط‌های اول پاراگراف اول، پاراگراف اول، خطوط بالای متن.

چگالی کلیدواژه (Weight/Density)

منظور از چگالی، نسبت دفعات تکرار کلمات کلیدی به کل متن است. باید مقدار این پارامتر را در حد معقول نگهدارید.

فرمول محاسبه درصد چگالی: $100 \times \text{تعداد کلمات متن} / \text{تعداد کلیدواژه‌ها}$

برای متن صفحه، چگالی بین ۳ تا ۷ مناسب است.

تعداد کلیدواژه (Frequency)

منظور از تعداد کلیدواژه، دفعاتی است که یک کلیدواژه بکار برده می‌شود. هرچه دفعات تکرار بیشتر باشد، نمره این پارامتر بالاتر می‌رود.

توزیع کلیدواژه (Proximity)

منظور این است که کلیدواژه‌ها چقدر از هم دور یا به هم نزدیک هستند. باید سعی شود که کلیدواژه‌ها خیلی به هم نزدیک نباشند و در همه جای متن وجود داشته باشند.

۴. تهیه نقشه سایت (Sitemap)

لیستی از تمام صفحات سایت به همراه توضیح کوتاهی در مورد آنها، در صفحه‌ای قرار دهید و لینکی به آن در صفحه اول سایت قرار دهید. نقشه سایت کار موتورهای جستجو را برای شناسایی صفحات سایت آسان می‌کند.

۵. محبوبیت لینک (Link Popularity)

این بخش از بهینه سازی، از اهمیت بسیار بالایی برخوردار است و از چند جهت بررسی می شود:

تعداد لینک های ورودی

اهمیت و کیفیت لینک های ورودی. لینکی که بین دو سایت با زمینه کاری مشابه ردوبدل شده، با کیفیت است و لینکی از یک سایت با رتبه بالا مهم است.

متنی که لینک می شود (Link Anchor Text). وجود کلیدواژه ای که صفحه بر مبنای آن بهینه سازی شده در این متن مهم است چون موتورهای جستجو این پارامتر را هم در نظر می گیرند.

معرفی به موتورهای جستجو و فهرست ها

پس از بهینه سازی موتورهای جستجو، باید آدرس و مشخصات سایت را در موتورهای جستجو ثبت کنید. معروف ترین موتورهای جستجو:

Google www.google.com/addurl

submit.search.yahoo.com/free/request : Yahoo

search.msn.com/docs/submit.aspx : MSN

فهرست های اینترنتی یا همان دایرکتوری ها جایی برای رده بندی سایت هاست که افراد حقیقی یا به عبارتی ادیتورها که انسان هستند آنها را کنترل می کنند و در نتیجه هر سایت بی خودی در آنها ثبت نمی شود. بنابراین ارزش لینک های داخل آنها بالاست و به همین خاطر برای موتورهای جستجو ارزش زیادی دارد. معروف ترین فهرست :

dmoz.org/World/Persian : ODP

موتورهای جستجوی بزرگ مثل یاهو و گوگل به طور منظم این دایرکتوری را چک میکنند و توجه خاصی به این دایرکتوری دارند

تبلیغات در موتورهای جستجو

پدیده جدیدی که در دنیای موتورهای جستجو به وجود آمده، تبلیغات مرتبط با متن است. به کمک این نوع از تبلیغات می توانید مخاطبین واقعی را هدف بگیرید و ترافیک سایت را بالاتر ببرید. ترافیک بالا عامل بسیار مهم دیگری است که بر رتبه سایت تاثیر می گذارد. استفاده از این نوع آگهی ها مزایای بسیاری دارد:

این آگهی ها برای کسانی به نمایش درمی آید که دقیقاً به دنبال خدمات یا کالاهای شما هستند.

مهم نیست آگهی ها چه تعداد نمایش داده شده باشند، شما تنها به ازای کلیک روی آنها، پول پرداخت می کنید.

می توانید سقف مبلغی که مایلید روزانه هزینه شود، تعیین کنید.

- می توانید کسانی که روی آگهی ها کلیک می کنند، ردیابی کنید.
- می توانید تعیین کنید آگهی ها در چه کشورهایی نمایش داده شوند.
- می توانید آگهی ها را به زبان های مختلف تهیه کنید.
- می توانید تعیین کنید آگهی ها در چه ساعاتی از روز نمایش داده شوند.
- می توانید بازدیدکننده ها را به صفحه کاملاً مرتبط با درخواستشان هدایت کنید.
- می توانید سرعت نمایش آگهی را تعیین کنید، یعنی به ازای هر چندبار جستجوی مرتبط، آگهی نمایش داده شود.
- از همان روزی که آگهی فعال شد، بیننده خواهید داشت.

فاکتورهای کاهش رتبه در موتورهای جستجو

- لینک های خراب (Broken Links)
- موتورهای جستجو اگر در بازدید از سایت متوجه لینکی معیوب شوند از بازدید سایر صفحات دست می کشند که این امر باعث کاهش رتبه سایت می شود .
- تقلب (Spam)
- چگالی بالای کلیدواژه، فاصله نزدیک آنها نسبت به هم، نوشته های بسیار کوچک و نوشته های هم رنگ بازمینه، به عنوان تقلب شناخته می شوند که موتور جستجو با آنها برخورد کرده و ممکن است سایت را به کلی از لیست خود حذف کند.
- صفحات آینه ای (Mirror Pages)
- منظور صفحات تکراری با آدرس متفاوت و محتوای یکسان است.
- مزرع لینک (Link Farms)
- مزرع لینک به گروهی از سایتها گفته می شود که بی آنکه ارتباط مفهومی باهم داشته باشند دست به مبادله گستره لینک می زنند تا محبوبیت خود را افزایش دهند. موتورهای جستجو اگر پی ببرند که شما در اینگونه سایتها قرار دارید، شدت با آن برخورد می کنند. نتیجه آن هم کاهش شدید جایگاه شما است.
- بررسی و آنالیز وبسایت
- آنالیز وبسایت نقش مهمی در بهبود کیفیت سایت و در نتیجه اثربخشی آن دارد. به موارد زیر توجه کنید:
- تعداد بازدیدکنندگان

✚ موقعیت جغرافیایی بازدیدکنندگان

✚ چگونگی ورود آنها به سایت (موتورهای جستجو، تایپ مستقیم آدرس شما، از سایتهای دیگر)

✚ با چه کلیدواژههایی سایت شما را پیدا کرده اند.

✚ روزها و ساعات بازدید

✚ تعداد صفحات بازدید شده

۹. سرویس جستجوی دایرکتوریاها

دایرکتوریاها (Directories) یک سلسله بانکهای اطلاعاتی هستند با لینکهایی به وبسایت مرجع، که این وبسایتها بوسیله اشخاص در زمان حال بوجود آمده است و بر اساس قانون مختص به سرویس جستجو طبقه بندی می شود. Yahoo مادر همه دایرکتوریاها است.

Look smart هم کاملاً عمومی است و این دایرکتوری را می توانید در سایتهایی مانند MSN پیدا کنید .

دایرکتوریاها بسیار مفید هستند. زمانیکه شما بیش از یک مفهوم یا توصیف برای آنچه که جستجو می کنید ندارید.

صفحه اول، معمولاً مقالات مورد نظر را به شما می دهد. اکنون از میان مقاله های موجود می توانید مقاله مورد نظر خود را بیابید و آدرس مقاله یا صفحه یا سایتی را که برای شما جالبتر است انتخاب کنید و شروع به خواندن کنید. اگر شما این شکل جستجو را برای پیدا کردن اطلاعات در زمینه مورد علاقه خود انتخاب کردید و به این روش سایتی را کشف کردید باید به خاطر داشته باشید که شما واقعاً آن متنی را که پیدا کرده اید جستجو نکرده اید، بلکه متنی را که شامل تیترا سایت و توصیف آن است جستجو کرده اید . اینها (تیترا و توصیف آن) بوسیله تنظیم کننده دایرکتوری و اغلب مبنی بر پیشنهاد صاحبان سایتها ساخته شده است. البته بعضی از دایرکتوریاها اگر نتوانند درخواست شما را پاسخگو باشند اطلاعاتی را از موتورهای جستجو برای شما پیدا می کنند.

تفاوت موتورهای جستجو با دایرکتوری ها (فهرستها)

موتورهای جستجو (مانند Alta Vista) از نظر توانایی در جمع آوری اطلاعات و مرتب سازی اتوماتیک آنها در خصوص موضوع مورد جستجو با فهرستها (مانند Yahoo) تفاوت دارند. دایرکتوری ها لیستهایی از پایگاه های اینترنتی هستند که بوسیله افراد تنظیم شده اند. نتیجه جستجوی شما بر روی یک موضوع قسمتی از این لیستهای تنظیم شده توسط انسان است که در ارتباط با موضوع مورد نظر شما است. با وجود این که مجموعه جواب یا نتیجه جستجوی به دست آمده در فهرستها (Directories) بسیار کوچکتر از نتیجه جستجو در موتورهای جستجو است اما جستجوی به دست آمده در دایرکتوری ها،



بیشتر مرتبط به موضوع می باشد و به همین دلیل از موتورهای جستجو بیشتر در پیدا کردن مطالب نادر و تخصصی و یا در تحقیقات جامع و کلی، استفاده می شود.



۱۰. خود آزمایی

۱. کدام یک از گزینه ها از بخشهای یک موتور جستجوگر می باشند:
الف- بخشهای Google, directory, crawler, spider
ب- بخشهای google, yahoo, Excite, Altavista
ج- بخشهای Indexer, Directory, Database, Ranker
د- بخشهای Ranker, Database, Indener, Spider, Crawler
۲. کدامیک از موتورهای جستجوی زیر از عملگرهای and, or, not استفاده می کنند؟
الف- موتورهای جستجو Excite, Altavista ب- موتورهای جستجو Yahoo, Altavista
ج- موتورهای جستجو Lycos, Google د- الف و ج
۳. لیست‌هایی از پایگاه‌های اینترنتی هستند که بوسیله افراد تنظیم شده‌اند.
الف. موتور جستجوی ترکیبی ب. دایرکتوری‌ها
ج. فهرستهای تکمیل دستی د. موتور جستجوی پیمایشی
۴. تکرار یا Frequency عامل بزرگ و مهم دیگری است که موتورهای جستجو از طریق آن صفحات مربوط را شناسایی می‌نمایند.
الف- صحیح ب- غلط
۵. هدف از صفحات وب جایگیری در صفحات نخست نتایج جستجو است.
الف- جستجو ب- بهینه سازی ج- ساخت د- رتبه بندی

۱۱. پاسخنامه

۱. د ۲. د ۳. ب ۴. الف ۵. ب



۱۲. کلیدواژگان

Directory	امکانات راهنمایی
Reported Size	اندازه گزارش شده
Partially-Indexed	ایندکس شده نسبی
Indexer	بایگانی کننده
Bot	برنامه های رهیاب
Database	پایگاه داده
Email	پست الکترونیکی
Link	پیوندهای
Frequency	تعداد
Spam	تقلب
Meta Description	تگ توصیف
Meta Keywords	تگ کلیدواژه ها
Proximity	توزیع کلیدواژه
Prominence	جایگاه
Search Box	جعبه جستجو
Weight/Density	چگالی
Crawling	خزش
Crawler	خزنده
Directories	دایرکتوریها
Categories	دسته بندی ها
Ranking	رتبه بندی
Robot	روبوت
Location/Frequency Methode	روش مکان - تکرار
Ranker	سیستم رتبه بندی



World Wide Web	شبکه جهانی وب
Mirror Pages	صفحات آینه‌ای
Web Pages	صفحات وب
Page Depth Amount	عمق صفحات
Spider	عنکبوت
Title	عنوان صفحه
Link-Only	فقط لینک
Human-Powered Directories	فهرستهای تکمیل دستی
News Group	گروه‌های خبری
Keyword	لغات کلیدی
Broken Links	لینک‌های خراب
Content	متن صفحه
Link Popularity	محبوبیت لینک
Link Farms	مزرعه لینک
Engines	موتورها
Search Engines	موتورهای جستجو
Crawler-Based Search Engines	موتورهای جستجوی پیمایشی
Sitemap	نقشه سایت



۱۳. منابع

۱. رودکی، مهدی. (۱۳۹۱). کتاب یک یا دوهزار و یک (کتاب سنو). چاپ سوم. نسخه ۳.۰ <http://www.iranseo.com>
۲. T.Joachims (۲۰۱۲). "Evaluation search engines using clickthrough data ". To appear.
۳. P.Faraday. (۲۰۰۱) "Attending to web pages". CHI ۲۰۰۱ extended abstracts (poster), pages ۱۰۰-۲۰۰
۴. Charles Knight.(۲۰۰۷). AltSearchEngines editor. The Top ۱۰۰ Alternative Search Engines, January ۲۸th. http://www.readwriteweb.com/archives/top_۱۰۰_alternative_search_engines.php
۵. Mike Grehan. (۲۰۰۲). How Search Engines Work Excerpted from the ۲۰۰۲ publication of Search Engine Marketing: The Essential Best Practice Guide. <http://www.mentevsual.com.ec/pdf/search-engines.pdf>
۶. Melissa S. Barker.(۲۰۱۰). .Top ۱۰ Internet Search Tips. *Illustrated ۸th Edition*, Course Technology. <http://www.melissabarker.com/SearchTips.pdf>

"پایان دوره آموزشی"