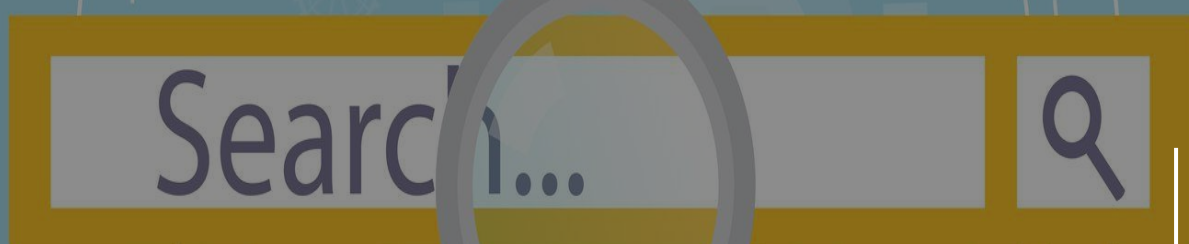




Namatek

True Education



Search Engine

www.namatek.com

آشنایی با موتور جستجو

فهرست مطالب

1. موتور جستجو چیست؟
2. موتور جستجو وب
3. موتور جستجو چگونه کار می کنند؟
4. اجزای مختلف موتور جستجو؟

با به کار بردن اصطلاح موتور جستجو (search Engine) مخاطب شما ناخودآگاه به یاد گوگل و یاهو می افتد. اما باید بدانید موتور جستجو تنها به معنی جستجو در گوگل و یاهو خلاصه نمی شود. در این مقاله سعی داریم به زبان ساده با این مفهوم آشنا شویم.



موتور جستجو چیست؟

در پاسخ به این سوال می توان گفت موتور جستجو بر اساس یک سری الگوریتم خاص اطلاعات مورد نیاز شما را از سایت های مختلف و همینطور پایگاه داده ها استخراج می کند و برای شما به نمایش می گذارد. به عبارتی دیگر به مجموعه ای از نرم افزارها و سخت افزارها که در یک فضای مشخص و در بین حجم عظیمی از داده اطلاعاتی که کاربر درخواست کرده است را جستجو می کند و به کاربر نمایش می دهد.

در عصر امروز استفاده از موتور جستجو به طرز چشم گیری افزایش یافته است. پس الگوریتم مورد استفاده به جهت قرار دادن اطلاعات درست در

دست کاربر بسیار مهم است. جستجو ها می تواند بر روی هارد دیسک، در یک پایگاه داده، در فضای وب و یا روی حافظه موبایل شما انجام گیرد.

موتور جستجو وب

در صورتی که موتور جستجو در وب به دنبال اطلاعات بگردد به آن موتور جستجو وب گفته می شود.

موتور جستجو چگونه کار می کند؟

کاربران بعد از سرچ تنها نتایج جستجوی خود را می بینند. اما این نتیجه حاصل کار بخش های متفاوت موتور جستجوگر است. در قدم اول بهتر است بدانید که موتور جستجوگر از قبل پایگاه داده خود را آماده کرده است و به این صورت نیست که درست در هنگامی که شما موضوع جستجو خود را تایپ می کنید، تمام سایت ها را برای بدست آوردن اطلاعات مورد نیاز شما بگردد.

این سوال در ذهن بسیاری از افراد وجود دارد که چگونه امکان دارد در کمتر از یک دقیقه موتور جستجو گر اطلاعاتی شامل میلیون ها سایت و صفحه را در اختیار کاربران خود قرار دهد. این کار نه تنها توسط گوگل بلکه توسط هیچ موتور جستجوگر دیگری قابل انجام نمی باشد. همه موتورهای جستجو بعد از جستجوی کاربر به پایگاه داده ای که از قبل در اختیار آنان قرار گرفته است سر میزنند و در آنجا شروع به جستجو می کنند. در نتیجه هیچ موتور جستجویی در وب جستجو نمی کند. موتور جستجوگر به کمک

قسمت های متفاوت خود اطلاعات را از قبل جمع آوری و سپس تجزیه و تحلیل می کند و در هنگام درخواست کاربران در همین پایگاه داده به دنبال اطلاعات خواسته شده می گردد.

اجزای مختلف موتور جستجو؟

عنکبوت (spider)

نرم افزار اسپایدر اطلاعات مورد نیاز یک موتور جستجوگر را جمع آوری می کند. اسپایدر با مراجعه به صفحات مختلف محتوای آن ها را بررسی می کند و اطلاعات مورد نیاز را جمع آوری می کند و در نهایت آن اطلاعات را در اختیار سایر بخش های موتور جستجوگر قرار می دهد. اسپایدر بسیار شبیه به کاربران وب عمل می کند. یعنی همانگونه که کاربران در هنگام جستجو از صفحات مختلفی بازدید می کنند اسپایدر هم دقیقا همین عمل را انجام می دهد.

تفاوت این دو نوع در اینجاست که اسپایدر کارهای قبل از سرچ را انجام داده و نتیجه حاصل را به کاربران نمایش می دهد. یعنی اسپایدر کدهای html صفحات را می بیند و این کدها را در کنار یکدیگر قرار می دهد و نتیجه صفحه موجود را به کاربر نشان می دهد.

خزنده (Crawler)

کراول به نرم افزاری می گویند که به عنوان یک فرمانده کار می کند و به اسپایدر دستور می دهد. کراول صفحات مورد بازدید اسپایدر را مشخص می کند و برای صفحه هایی که اسپایدر در حال حاضر در آن ها قرار دارد تصمیم گیری می کند و مشخص می کند که دنبال شوند و یا خیر. حتی ممکن است که کراول، از قبل طبق برنامه ریزی قبلی یک سری آدرس های خاص را در اختیار اسپایدر قرار دهد.

حجم اطلاعات ذخیره شده در پایگاه داده موتور جستجوگر به دنبال کردن لینک های یک صفحه ارتباط مستقیم دارد. حتی امکان این موضوع وجود دارد که دسترسی به برخی صفحات به موتور جستجوگر داده نشده باشد. موتور جستجو قبل از ورود به هر سایتی ابتدا قوانین آن سایت را مطالعه می کند. شما به عنوان یک صاحب سایت می توانید به موتور جستجو اجازه دسترسی به سایتتان را بدهید و یا این دسترسی را از آن بگیرید. تنظیم اجازه دسترسی موتور جستجو به یک سایت توسط پروتکل robots انجام می پذیرد. شما می توانید با مراجعه به آن قسمت این کار را انجام دهید.

بایگانی کننده (Indexer)

تمامی اطلاعاتی که توسط اسپایدر جمع آوری شده است در اختیار ایندکسر قرار می گیرد. در این بخش اطلاعات ارسالی را مورد تجزیه و تحلیل قرار می دهند و آن اطلاعات را به بخش های متفاوت دیگری تقسیم بندی می کنند. نتیجه تجزیه و تحلیل بدین شکل است که صفحه ای که اطلاعات از آنجا

ارسال شده اند، حجم متن، کلمات موجود، میزان تکرار کلمات، جایگاه قرار گیری کلمات و.... در آن مشخص می شود.

در واقع کار ایندکسر، خورد کردن پارامترهای صفحه است و سپس تبدیل این پارامترها به یک مقیاس عددی تا اینکه سیستم رتبه بندی توانایی مقایسه صفحات مختلف با یکدیگر را داشته باشد. به جهت پایین آوردن حجم داده ها ایندکسر از تجزیه کلمات رایج صرف نظر می کند.



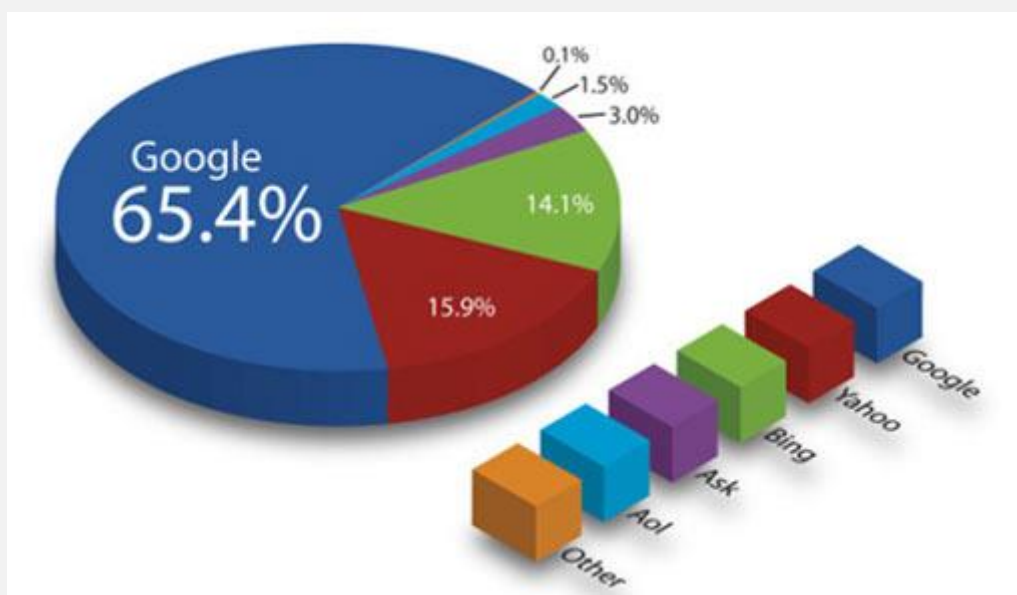
پایگاه داده (Database)

در ایندکسر تمامی داده هایی که تجزیه و تحلیل می شوند به پایگاه داده ارسال می شوند. در این قسمت داده ها گروه بندی، کدگذاری و با تکنیک های خاصی ابتدا فشرده و سپس ذخیره می شوند. این فشرده سازی تنها به دلیل کم کردن حجم اشغالی انجام می شود. یک موتور جستجوگر لازم است که پایگاه داده عظیمی داشته باشد و اطلاعات قدیمی خود را دائماً آپدیت کند و دائماً حجم محتوای خود را افزایش دهد.

اینکه یک موتور جستجوگر دارای پایگاه داده ای بزرگ و همچنین به روز باشد برای آن پایگاه داده یک امتیاز مثبت حساب می شود. از تفاوت های اصلی موجود در موتورهای جستجوگر تفاوت در حجم پایگاه داده و همینطور روش ذخیره سازی اطلاعات است.

سیستم رتبه بندی (Ranker)

بلافاصله بعد از انجام پذیری مراحل قبلی موتور جستجوگر آماده پاسخگویی به سوالات کاربران خود است. کافی است کاربران در قسمت جستجو کلمه و یا موضوع مورد نظر خودشان را تایپ کنند و سپس بر روی آیکن جستجو و یا دکمه enter ضربه بزنند و منتظر پاسخ بمانند.



برای اینکه نتیجه ای که شما خواسته اید به شما نمایش داده شود موتور جستجو در ابتدا تمامی صفحات موجود در پایگاه داده که به موضوع مطرح شده از شما مرتبط هستند را مورد مطالعه و بررسی قرار می دهد و سپس سیستم رتبه بندی وارد عمل می شود و نتایج را از بیشترین ارتباطی که با

موضوع درخواستی شما دارند تا کمترین ارتباط به ترتیب مشخص می کند و سپس نمایش می دهد.

اگر موتور جستجوگر نتواند پاسخ درست و مرتبطی با موضوع درخواستی شما ارائه دهد اگرچه بهترین و حتی کارآمدترین پایگاه داده را داشته باشد، یک موتور جستجو ضعیف است. سیستم رتبه بندی به عنوان قلب تپنده موتور جستجوگر عمل می کند. سیستم رتبه بندی برای اینکه بهترین و کارآمدترین پاسخ را در اختیار کاربران خود قرار دهد پارامترهای زیادی را در نظر می گیرد. تا به کنون از سیستم رتبه بندی گوگل قدرتمندتر وجود نداشته و گوگل در این عرصه هنوز هم پیشتاز است.