

تهدیدها، فرصت ها و ملاحظات موتور جستجوی اینترنت



فهرست

۳ مقدمه
۵ فصل اول: معماری موتورهای جستجو
۱۳ فصل دوم: موتورهای جستجوی ملی و اهمیت آنها
۱۸ فصل سوم: پدافند غی‌عامل و سامانه موتور جستجو

در حال حاضر میلیاردها صفحه وب، حاوی اطلاعات فراوانی در حوزه‌های متفاوت روی کارگزار^۱های مختلف در سطح جهان قابل دسترسی هستند. نرخ رشد اطلاعات به گونه‌ای است که امروزه موضوع نحوه دسترسی، مدیریت و دسته بندی اطلاعات بسیار جدیتر از مقوله جمع آوری اطلاعات است.

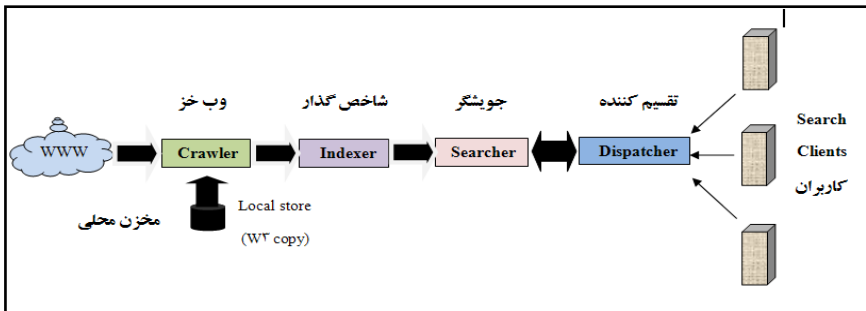
رشد روزافزون کارکرد اینترنت در فضای تبادل اطلاعات و نقش انکارناپذیر موتورهای جستجو در راهبری کاربران فضای مجازی به مقصد مورد نظر، اهمیت پرداختن به آن‌ها را بوضوح تبیین می‌کند. موتور جستجو، سامانه‌ای است که به منظور یاری کاربران در جهت یافتن اطلاعات مد نظر در بین بی‌شمار صفحه اینترنتی طراحی شده است. امروزه ابزارهای کاوش و تحلیل اطلاعات، همچون موتورهای جستجو از جایگاه مهمی در ایجاد امکان برآوردن نیازهای اطلاعاتی کاربران برخوردار هستند. مواردی همچون ضرورت حفظ امنیت اطلاعات در مواجهه با رفتار موتورهای جستجو سبب می‌شود تا ایده ایجاد موتور جستجوی ملی به صورت جدی برای هر کشور و ملت خواهان حفظ استقلال مطرح باشد. لذا در راستای برقراری امنیت فضای تبادل اطلاعات، ضرورت ایجاد موتور جستجوی بومی که ملاحظات پدافند غیر عامل در طراحی و ساخت آن رعایت شده باشد احساس می‌شود.

¹ Server

فصل ۱

معماری موتورهای جستجو

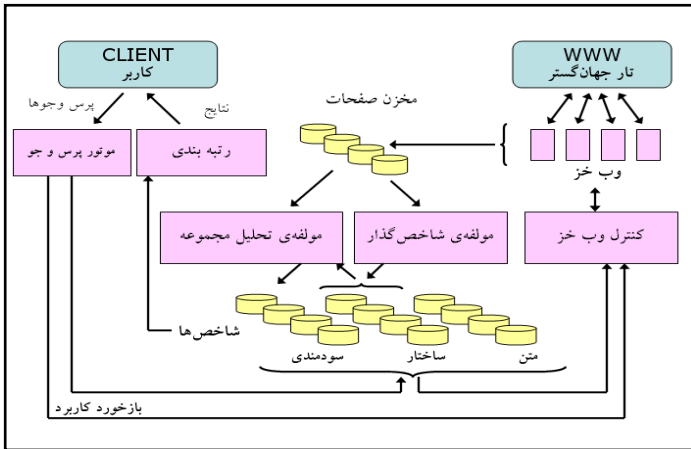
در این بخش معماری کلی موتورهای جستجوی اینترنت را به طور اجمالی مورد بررسی قرار می‌دهیم.



مدل ۱. شمای سطح بالای موتور جستجو

مروری بر ساختار کلی یک موتور جستجو

اکثر موتورهای جستجو از مؤلفه های وب خز^۱، شاخصگذار^۲، جویشر^۳ و تقسیم کننده امکانات^۴ تشکیل شده اند. نمونه این معماری در سامانه هایی مانند Google، Altavista و FAST دیده می شود.



مدل ۲. ساختار کلی، اجزا و ارتباطات موتور جستجو

همان طور که در شکل فوق دیده می شود مؤلفه کنترل وب خز مسئول مسیریابی به عملیات وب خز است؛ مؤلفه شاخص گذار تمامی پیوندهای صفحات را بیرون کشیده و URL ای که هر یک از صفحات در آن دیده شده اند را ثبت می کند و مؤلفه تحلیل مجموعه^۶ مسئول ایجاد شاخص های متنوع دیگر است.

¹ Crawler

² Indexer

³ Searcher

⁴ Dispatcher

⁵ Link

⁶ Collection Analysis Module

شاخص سودمندی^۱ در مدل ۲ توسط مؤلفه تحلیل مجموعه ایجاد شده است. برای مثال این شاخص امکان دسترسی به صفحات حاوی مطالب مهم و یا حاوی بیش از چند عکس را می‌دهد. مخزن صفحات^۲ در زمان وب خزی و شاخص گذاری، صفحاتی را که از وب می‌گیرد در پایگاه مربوطه ذخیره می‌کند. مؤلفه موتور پرس و جو مسئول دریافت درخواست های کاربران است. با توجه به حجم اطلاعاتی وب و اینکه کاربران معمولاً یک یا دو لغت را جستجو می‌کنند، مؤلفه رتبه‌بندی^۳ نقش مهمی را در انجام یک جستجوی مطلوب ایفا می‌کند. در ادامه برای هر یک از اجزای معماری تعریف مختصری آورده شده است.

وب خز

مؤلفه جمع‌آوری مستندات از وب است. جهت جمع‌آوری مستندات چندین روش ابتکاری و الگوریتمی برای وب خزها وجود دارد که اکثر آنها مبتنی بر تعقیب پیوندها در اسناد ابر متنی^۴ هستند. یک وب خز^۵ با نام‌های مختلفی مانند عنکبوت^۶، ربات^۷، حشره^۸ و شاخصگذار خودکار شناخته می‌شود. وب خزها یک کپی از صفحات مرور شده را ایجاد و ذخیره می‌کنند و هنگامی که صفحات را مرور می‌کنند، ابر پیوندهای هر صفحه را شناسایی کرده و آن‌ها را به فهرست خود اضافه می‌کنند.

¹ Utility index

² Page Repository

³ Ranking

⁴ Hyper-text

⁵ Crawlers

⁶ Spider

⁷ Robot

⁸ Ant

دام عنكبوت یا دام وب خزی یکی از آسیب‌پذیریهای این مؤلفه است که شامل مجموعه‌ای از صفحات وب میشود که به صورت عمدی یا غیر عمدی وب خزی را با بی‌نهایت درخواست مواجه می‌کنند. روش‌های عمومی دام عنكبوت عبارتند از:

- استفاده از صفحات پویا مانند تقویم که تعداد بی‌نهایتی صفحه ایجاد می‌کنند که وب خزی باید آنها را دنبال کند.
- استفاده از صفحاتی که با تعداد زیادی کاراکتر پر شده‌اند که تحلیل‌گر لغوی صفحه را با مشکل روبرو می‌کند.

✚ شاخصگذار

مؤلفهای است که مجموعه‌ای از اسناد و داده‌ها را اخذ و شاخص قابل جستجو برای آنها می‌سازد.

✚ جویشگر

این مؤلفه از یک سو فایل خروجی شاخصگذار را بعنوان ورودی پذیرفته و از سوی دیگر پرسش کاربران را از طریق تقسیم‌کننده دریافت می‌کند. جویشگر پرسش کاربر را بر روی قسمت‌هایی از شاخص‌هایش اجرا می‌کند و نتایج مرتب شده از جستجو را همراه با شناسه سند و امتیاز مرتبط^۱ به تقسیم‌کننده برمی‌گرداند.

¹ Related Score

✚ تقسیم کننده

تقسیم کننده پس از دریافت پرسش کاربر، از فهرست جویشرها یکی را برای اجرا انتخاب کرده و پرسش را برای آن می فرستد. در نهایت نیز فهرست مرتب نتایج را از جویشرها میگیرد. تقسیم کننده برای هر نتیجه، یک شناسه سند منحصر به فرد و یک امتیاز مرتبط دریافت می کند. فهرست های جویشرها با هم ادغام می شوند تا نتایج مرتبط با امتیازهای بالاتر به کاربر نمایش داده شوند.

واسط کاربری موتور جستجو

واسط پرسش اصلی ترین بخش واسط کاربری موتور جستجو بشمار می آید. این واسط در واقع صفحه ای است که کاربران هنگام ورود به موتور جستجو آن را مشاهده کرده و واژگان مورد نظر خود را وارد می کنند. اجزای اصلی این واسط بطور کلی شامل یک کادر متنی و یک دکمه برای فعال سازی آن است. رعایت مراحل چارچوب چهار مرحله ای ذیل باعث طراحی یک واسط کاربری مناسب و استاندارد با ویژگی های متنوع در قالبی سازگار خواهد شد.

✚ فرمول سازی^۱: بررسی رخدادهای قبل از شروع جستجو

✚ عملیات^۲: چگونگی شروع جستجو

✚ بازیابی نتایج: بررسی چستی و چگونگی نمایش نتایج جستجو

✚ پالایش^۳: بررسی اتفاقات بعد از بازیابی نتایج و پیش از بازگشت دوباره به

مرحله فرمول سازی

¹ Formulation

² Action

³ Refinement

استانداردهای کاربردی موتورهای جستجو

موتورهای جستجو عموماً برای دریافت پرسش کاربر از پروتکل HTTP استفاده میکنند. استانداردهای مختلفی در این زمینه ارائه شده است. ایده اصلی پشت همه این استانداردها این است که هر موتور جستجویی باید بتواند یک فرمت ورودی استاندارد را قبول کند. در ادامه بعضی از مهمترین استانداردها در این حوزه بیان میشود.

استاندارد Z39.50

Z39.50 یک استاندارد بینالمللی برای ارتباط بین دو سامانه رایانهای به خصوص، کتابخانه و سامانههای مرتبط با اطلاعات است. این استاندارد توسط ISO و ANSI/NISO پشتیبانی میشود.

استاندارد SRU

SRU¹ مبتنی بر REST² بوده و امکان توصیف پرسشها به صورت رشتههای پرسشی URL و به زبان CQL³ را میدهد؛ SRU از SOAP⁴ استفاده کرده و در اصل مبتنی بر پروتکل HTTP است.

استاندارد OpenSearch

OpenSearch مجموعه‌ای از فرمت‌های ساده برای به اشتراک‌گذاری نتایج جستجو است و شامل تعاریف XML برای توصیف قابلیت‌های موتور جستجو، فرمت خروجی و یک زبان پرس و جوی ساده به شکل URL می‌شود.

¹ Search/Retrieval via URL

² Representational State Transfer

³ Common Query Language

⁴ Simple Object Access Protocol

✚ استاندارد منع ربات‌ها

منع ربات‌ها^۱ قراردادی است که موتورهای جستجو را از دسترسی به همه یا بخشی از یک وب سایت منع میکند. با توجه به این که فهرست تمامی صفحاتی که نمیخواهیم شاخص‌گذاری شود را نوشته و در اختیار عموم قرار داده‌ایم، استفاده غلط از این استاندارد می‌تواند باعث شود به صورت ناخواسته برخی اطلاعات را در اختیار مهاجمان قرار گیرد.

✚ استاندارد ACAP^۲

ACAP در خصوص نحوه شاخص‌گذاری موتورهای جستجو با همکاری ناشران و موتورهای جستجو برای تغییر روال تولید، انتشار و استفاده از محتواهای License-protected روی اینترنت در سال ۲۰۰۷ ارائه شده است.

آینده موتورهای جستجو

از جمله ایده‌های جدید در حوزه موتورهای جستجو میتوان به موارد زیر اشاره کرد:

✚ ارتباط دوطرفه و تعامل با کاربر

✚ توسعه خوشه‌بندی نتایج قبل یا بعد از عمل جستجو

✚ موتورهای جستجوی چندرسانه‌ای با روش پرسوجو در بین قالبهای

مختلف با قابلیت پردازش گفتار و تصویر

^۱ Robots Exclusion Standard

^۲ Automated Content Access Protocol

✚ موتورهای جستجوی معنایی که تنها به شاخص کلمات وابسته نیستند و می‌توانند محتوای صفحات وب را درک کرده و بر روی آنها استنتاجات منطقی انجام دهند.

✚ موتورهای جستجوی اجتماعی¹ که نوعی از موتورهای جستجو است که میزان ارتباط نتایج جستجو را با توجه به قضاوت یا همکاری کاربران بدست می‌آورد.

¹ Social search engine

فصل ۲

موتورهای جستجوی ملی و اهمیت آنها

در حال حاضر بیش از ۲۶۰۰ موتور جستجوی سفارشی شده در ۲۱۶ کشور جهان وجود دارد. موتور جستجوی سفارشی شده میتواند یک موتور جستجوی ملی، منطقه‌ای یا مربوط به یک زبان خاص باشد. ویژگی اصلی این موتورهای جستجو توجه ویژه محتوا یا کاربران به یک حوزه سازمانی، منطقهای، فرقه‌ای، زبانی و مواردی از این قبیل است. امروزه موتورهای جستجو در شاخه های مختلف تخصصی نظیر حسابداری، کسب و کار، بنگاهی و سازمانی، پزشکی و غیره در حال ارائه سرویس به کاربران خود هستند. در این بخش به چند نمونه موفق از موتور جستجوهای ملی از مجموعه فوق، اشاره می‌کنیم.

چین

یکی از موفقترین موتور جستجوهای ملی، موتور جستجوی Baidu در چین است. این موتور جستجو می تواند وبسایتها، فایلها، تصاویر و تصاویر را کاوش کند. بطور کامل زبان چینی را پشتیبانی کرده و سرویس های مختلفی نظیر جستجوی کتاب و واسط کاربری مناسب ارائه می کند.

روسیه

موتور جستجو و پورتال Yandex از سال ۱۹۹۷ فعالیت خود را شروع کرده است. مزیت مهم این موتور جستجو پشتیبانی از خصوصیات زبان روسی است و پایگاه داده بزرگ آن در روسیه است.

هند

هدف موتور جستجو Guruji، محصول سال ۲۰۰۶، تمرکز بر محتوای هندی است. این موتور، امکان جستجو در وب سایتها، فیلمها و کسبوکار را برحسب محل انجام می دهد و زبانها و خطوط مختلف هندی را پشتیبانی می کند.

عربستان

موتور جستجوی Araby از سال ۲۰۰۶ فعالیت خود را آغاز کرد. این موتور جستجو علاوه بر جستجوی وبسایتها، عکسها و ویدئوها، جستجوی تخصصی در موضوعات مختلف نظیر اسلام را نیز امکان پذیر می سازد.

اهمیت ایجاد و استفاده از موتورهای جستجوی ملی

از منظرهای گوناگون می توان به بررسی اهمیت موتورهای جستجوی ملی پرداخت:

استقلال در پویش اطلاعات شبکه داخلی

با ایجاد شبکه اینترنت ملی میتوان محتوای اطلاعاتی داخل کشور را بدون نیاز به شبکههای خارجی ذخیره و احصا کرد، ولی این در صورتی است که کاربر آدرس مکانی که مطلب مورد نظر ذخیره شده است را بداند. لذا نیاز به موتور جستجویی که بدون نیاز به سرویسهای خارجی و موتور جستجوهای دیگر بتواند این اطلاعات را در اختیار کاربران قرار دهد احساس می شود.

ویژگی های خطی و زبانی

با استفاده از موتور جستجوهای سفارشی شده می توان از قابلیت های خطی و زبانی مربوطه جهت بهبود نتایج فرآیند جستجو استفاده کرد. این موارد شامل سهولت استفاده کاربران و بالا رفتن دقت نتایج پرس و جو است.

حفاظت از آثار ادبی، فرهنگی و دینی

در بعضی از حوزهها نظیر مسائل دینی و ملیتی، عملکرد غیرمنصفانه موتورهای جستجو در ارائه اطلاعات به کاربران کاملاً مشهود است. حفظ آثار ادبی کهن و کلاسیک یکی دیگر از سرویسهایی است که می تواند انگیزه ساخت یک موتور جستجوی سفارشی شده، باشد. از سوی دیگر با استفاده از یک موتور جستجوی سفارشی شده میتوان نتایج جستجو را با توجه به جهت دهی های فرهنگی - ملی تعیین کرد.

✚ شناخت گرایش های جامعه و مدیریت آن

با استفاده از موتورهای جستجوی سفارشی شده مدیران و برنامه ریزان میتوانند با توجه به درخواستهای ارسال شده کاربران، بسته به مکان و سابقه جستجوگران، بازخوردهایی از وضعیت کنونی جامعه دریافت کنند.

✚ ابعاد اقتصادی

با توجه به اهمیت تجاری موتورهای جستجو و منابع درآمدی آن ها از محل هایی نظیر فروش تبلیغات و ارائه سرویسهای جانبی به کاربران، در کشورهای مختلف از موتورهای جستجو برای حمایت از کسبوکار محلی خود استفاده شده است.

چند نمونه از موتور جستجوهای داخلی

در این قسمت جهت آشنایی با وضع فعلی کشور در این حوزه چند نمونه از موتورهای جستجوی داخلی را معرفی و مختصراً بررسی میکنیم.

✚ بگردیم

این موتور جستجو سعی کرده است مشکلات موجود در دیگر موتورهای جستجو نظیر گوگل در مورد متون فارسی را حل کند. از جمله این امکانات می توان به پشتیبانی از عملگرهای فارسی مثل "و" و "یا"، شناسایی کاراکترهای مختلف برای نمایش یک حرف، و حذف کلمات مشابه فارسی اشاره کرد.

پارسیک

سایت پارسیک در اردیبهشت ۱۳۸۱ و با هدف ایجاد خدمات جستجو به
فلوسی‌زبانان تأسیس شده است. جستجوگر پارسیک یک ابر موتور جستجو^۱
است.

ریسمون

این موتور جستجو شاخصه‌هایش را خود می‌سازد و مبتنی بر موتور جستجوی
دیگری نیست. ریسمون از سایت Link.ir برای شناسایی سایت‌های مورد
بررسی استفاده می‌کند.

^۱ Meta Search Engine نوعی موتور جستجو است که درخواست کاربران را به چندین موتور
جستجو یا پایگاه داده فرستاده و پس از جمع‌آوری نتایج، آنها را تحت یک فهرست یا بر اساس
منابع‌شان نمایش می‌دهد.

پدافند غیر عامل و سامانه موتور جستجوی اینترنت

موتورهای جستجو نقش بسزایی در راهنمایی کاربران به محتوای مورد نظر و تکمیل روند ارائه کالا، خدمات و محتوا به صورت اینترنتی ایفا می کنند. قطع یا تحریم سرویس موتورهای جستجو عملاً دسترسی کاربران داخلی به سایت‌های داخلی و خارجی و دسترسی سایت‌های داخلی به کاربران داخلی و خارجی را قطع خواهد کرد. از سوی دیگر با توجه به کارکرد این سامانه ها روزانه داده‌های فراوانی از نحوه جستجوی افراد، نیازها و تمایلات آنها در اختیار سرویس دهندگان قرار می گیرد. این موضوع ضرورت طراحی و پیاده سازی موتورهای جستجوی بومی و ملی کارا، مطمئن و پایدار را نشان می دهد، در غیر این صورت انباشت ذیقیمتی از اطلاعات مربوط به فضای داخلی، طرز فکر و تلقی مردم در موضوعات مختلف و مواردی از این دست به راحتی در اختیار بیگانگان قرار می گیرد.

با توجه به شرایط حاضر، نکات ذیل جهت حفظ حریم شخصی کاربران در هنگام استفاده از موتورهای جستجو ارائه می شود:

✚ عدم استفاده از سرویسهای حساب کاربری در زمان جستجو

✚ عدم استفاده از سرویسهای جستجو ISP

✚ تغییر IP یا استفاده از برنامههای مخفی کننده IP

✚ جلوگیری از ثبت Cookieها از طرف موتور جستجو

ملاحظات پدافند غیرعامل در طراحی موتور جستجوی ملی

نکته مهم در طراحی یک موتور جستجوی داخلی علاوه بر کارایی و پایداری، موفقیت در جذب کاربران است. نمونه‌های مشابه نشان داده است که موتورهای جستجوی منطقه ای تنها زمانی توانسته‌اند با نمونه‌های مطرح جهانی رقابت کنند که از مزیت‌های خاص منطقه ای نظیر زبان، کتاب‌ها و خدمات محلی بهره کافی برده باشند. در ادامه برخی از مهمترین ملاحظات پدافند غیرعامل در طراحی موتورهای جستجو در کشور بیان شده است.

✚ از فناوری‌های غیربومی که کد آن قابل دسترس نیست استفاده نشود. همچنین استفاده از فناوری‌های متن باز به تنهایی کافی نیست بلکه باید از روش‌های بهینه‌سازی کدها و اضافه کردن بسته‌ها و سرویس‌های امنیتی سخت افزاری یا نرم‌افزاری بومی به اجزا، بستر معماری و ارتباطی آن نیز استفاده کرد.

✚ به منظور حفظ امنیت اطلاعات تبادل شده بین اجزای موتور جستجو باید از استانداردها و پروتکل‌های امنیتی بومی استفاده شود. در غیراین صورت مهاجم بین راه می‌تواند داده‌های مورد جستجو را شنود کرده و به تحلیل جستجوی افراد دست یابد. همچنین می‌تواند در نتایج جستجوی افراد تأثیر گذاشته و داده‌های تبادل شده را تحریف کند. پروتکل استفاده شده می‌تواند از الگوریتم‌های رمزنگاری داخلی و خاص سازمان استفاده کند.

✚ مکانیزم‌های مناسب برای حفظ الگوریتم‌های استفاده شده و محرمانه ماندن معماری نرم‌افزار اعمال شود. مهاجمان و هرزنگارها با داشتن اطلاعات دقیق از الگوریتم‌ها و معماری موتور جستجو راحت‌تر می‌توانند به آن حمله کرده و آن را از کار بیاندازند. لذا این نکته نه تنها

باید در طراحی نرم‌افزار مورد توجه قرار گیرد بلکه تیم طراحی و پشتیبانی آن نیز باید آموزش‌های لازم را دیده باشند و در حفظ مسائل امنیتی آن حداکثر تلاش خود را بکنند.



✚ ملاحظات پدافند غیرعامل در بسترهای نرم‌افزاری نظیر پایگاه داده، مستندات، سیستم عامل و کارگزار وب باید رعایت شود. از آنجاییکه کارکرد موتور جستجو وابسته به سامانه های نرم‌افزاری و سخت‌افزاری بستر آن است، عدم رعایت این ملاحظات امنیت و پایداری موتور جستجو را با خطر مواجه می‌کند.

✚ موتور جستجو برای ارائه سرویس خود نباید به کارگزارهای خارج کشور و تنها به محتوای خارجی وابسته باشد. بدین منظور ماشین‌های کارگزار موتور جستجو باید در داخل کشور باشند. همچنین با توجه به پیاده‌سازی شبکه اینترنت ملی، خدمت‌دهی موتور جستجو نباید وابسته به شبکه ارتباطی خارج از شبکه داخلی باشد. این خود امکان خروج ترافیک داده از مرزهای کشور را کاهش داده و می‌تواند باعث افزایش امنیت داده‌ها شود.

✚ معماری موتور جستجو باید قابلیت توزیع شدن و تکرارپذیری در مؤلفه‌هایی مانند وب‌خز، شاخص‌گذار و سامانه ذخیره‌سازی را داشته باشد. تکرارپذیری اجزای معماری جهت افزایش کارایی، جلوگیری از اشباع شدن و افزایش پایداری سامانه است. همچنین با توزیع کردن داده‌ها در کارگزارهای مختلف پایداری سامانه افزایش می‌یابد. آنچه که در هنگام توزیع‌پذیری از دیدگاه امنیتی مهم به نظر می‌رسد حفظ یکپارچگی و جامعیت در هنگام توزیع‌کردن فعالیت‌ها و وظایف است تا بتوان تعادلی بین سرعت و دقت و امنیت ایجاد کرد. موتورهای جستجو حتی با داشتن بخشی از داده‌های خود نیز باید بتوانند به پرسش‌های کاربر پاسخ دهند.

✚ در هنگام موازی‌سازی و تکرار، مؤلفه‌های موتور جستجو باید در مناطق جغرافیایی مختلف پراکنده شود. این عمل در کنار استقلال مؤلفه‌ها و مقاومت در برابر خرابی یک یا چندین جزء موتور جستجو نقش بسزایی در افزایش پایداری سامانه خصوصاً در جنگ‌های فیزیکی دارد.

✚ ردگیری و تحلیل پرسشها و رفتار کاربر در موتور جستجو تنها برای افراد مجاز ممکن باشد. اطلاعات آماری و تحلیلی در مورد جستجوهای دسته‌های مختلف کاربران اطلاعات مهمی هستند که ارزش سیاسی، اجتماعی و اقتصادی بالایی دارند. دسترسی افراد بدخواه به این اطلاعات می‌تواند آن‌ها را در برنامه‌ریزی برای ایجاد تنش‌های اجتماعی، سیاسی و اقتصادی در کشور یاری کند. از طرف دیگر این داده‌ها می‌تواند برای دولت مردان و برنامه‌ریزان کشور کمک بسزایی باشد. لذا دولت‌ها در کشورهای مختلف محدودیت‌های خاصی روی نحوه انتشار این اطلاعات وضع می‌کنند.

✚ مکانیزم بازنمایی موتور جستجو باید کاربر را از خطرهای امنیتی موجود در نتایج جستجو مطلع و حفاظت کند. گزارش‌های بسیاری وجود دارد که موتور جستجویی در هنگام نمایش نتایج جستجو با نمایش بخشی از محتوای ناامن سایت موجب به خطر افتادن امنیت کاربران خود شده است. موتور جستجو باید این گونه محتوا را در هنگام نمایش نتایج جستجو حذف کند. همچنین بعضی مهاجمان با استفاده از موتور جستجو و کلیدواژه‌های جذاب، کاربران را به سایت خود می‌کشند. موتور جستجو باید خطر سایت‌های ناامن را به کاربرانی که روی لینک آنها کلیک می‌کنند هشدار دهد. با این کار از حمله فرد مهاجم به وسیله موتور جستجو جلوگیری می‌شود.

✚ الگوریتم جمع‌آوری و رتبه‌بندی موتور جستجو باید تدابیر لازم برای جلوگیری از هرزنگاری را در نظر گرفته باشد. شخص مهاجم با استفاده از هرزنگاری می‌تواند سایت‌های مورد نظر خود را در صدر نتایج جستجوی کاربران قرار دهد. با این کار عملاً موتور جستجو از ارائه سرویس باز مانده و تبدیل به وسیله‌ای تبلیغاتی برای مهاجمان می‌گردد.

✚ اعمال درست سیاست‌های وب‌خز در موتور جستجو به منظور بالا بردن پایداری و کارایی موتور جستجو و مقابله با دام‌های وب‌خز. این سیاست‌ها باید با توجه به محتوا و کاربران هدف موتور جستجو به گونه‌ای انتخاب شود که علاوه بر حفظ کارایی و سودمندی موتور جستجو پایداری آن را خدشه‌دار نکند. مثلاً دام وب‌خز می‌تواند آن را در حلقه‌ای بی‌نهایت به دام انداخته یا منجر به اتلاف فضای زیادی از مخزن ذخیره‌سازی گشته و پایداری سامانه را دچار خدشه کند.

✚ موتور جستجو باید قابلیت پشتیبانی از زبان فارسی را در واسط کاربری، جستجو و بازنمایی داشته باشد. همچنین واسط‌های کاربری غیربومی باعث وابستگی افراد جامعه به کاربری این گونه واسط‌ها شده و آنها را از زبان بومی دور نگه می‌دارد.

✚ در جمع‌آوری اطلاعات و همچنین ثبت تعامل کاربران با سامانه باید مسائل مربوط به حریم شخصی کاربران رعایت شود. در واقع خود موتور جستجو نباید ناقض حریم امنیتی کاربران خود باشد. این نقض حریم می‌تواند در تحلیل اطلاعات تعامل کاربر با سامانه یا دسته‌بندی و ارائه داده‌ها به شکلی باشد که حمله فرد مهاجم (خصوصاً حملات مربوط به مهندسی اجتماعی) را تسهیل کند. به این ترتیب فرد مهاجم از موتور جستجو برای یافتن سریع‌تر هدف خود و حمله به آن بدون ردپا استفاده می‌کند. هر سازمان یا ارگانی که اطلاعات خود را در وب قرار می‌دهد باید روالی تکراری برای بررسی محتوای اطلاعاتی که از سازمان در اختیار عموم قرار می‌گیرد داشته باشد و این اطلاعات را قبل از فرد مهاجم یافته و از روی وب حذف کند.

✚ در جمع‌آوری اطلاعات موتور جستجو و بازنمایی آن‌ها به کاربر باید موارد حقوق مؤلفین در نظر گرفته شود. موتور جستجو باید حقوق مؤلفین را طبق یک پروتکل حقوق محتوای دیجیتال توافقی در مورد ارائه خدماتی مثل جستجوی کتاب‌ها یا محتواهای غیر رایگانی که سایت‌های خبری، کتابخانه‌ها و آرشیوها تنها جهت جستجو شدن در اختیار موتور جستجو قرار داده‌اند رعایت کند.