**مبانی جست و جو در «اینترنت»**

**قسمت اول**

**تهیه کننده:‌ محمد رضا معبودیان اصفهانی maboodian@isu.ac.ir**

به رغم افق نوینی که اینترنت و امکانات گوناگون آن به ویژه WEB در دسترسی یکپارچه و جهانی به اطلاعات، فرا روی انسان نهاده است، مشکلات بسیاری در جست و جو (search) و بازیابی اطلاعات به همراه دارد که از میزان موفقیت در دسترسی به اطلاعات مناسب و دلخواه می‎کاهد. زیرا برخلاف اطلاعات موجود در کتابخانه‎ها و پایگاه‎های اطلاعاتی که با توجه به معیارهای مشخص و استاندارد شده، منابع اطلاعاتی را برای افزودن به مجموعه خود و یا نمایه‎سازی انتخاب می‎کنند، منابع موجود در اینترنت را افراد و سازمان‎های گوناگون و با هدف‎های متفاوت به آن می‎افزایند. به عبارت دیگر اکثر منابع اینترنتی،‌به ناچار بدون ضابطه والگوی مشخصی وارد اینترنت می‎شود.

ویلیام کنز می‎نویسد«اینترنت همچون مغازه سمساری است که همه چیز در آن یافت می‎شود. اینترنت پدیده‎ای بی‎شکل و قواره و در عین حال بسیار گرسنه است. بنابر این یک فرد بی‎تجربه ممکن است ساعتهای مدیدی غرق در جست و جوی بی‎فایده در اینترنت شود. حتی جست و جوهایی که با منطق اولیه شروع می‎شوند ممکن است بعدها در فضای نامتناهی شبکه گم شوند». و بی‎نتیجه بمانند!

در نگاه اول بازیابی اطلاعات در اینترنت امری بسیار ساده است: جوینده اطلاعات با وارد کردن آدرس سایت در مرورگر اینترنت، به اطلاعات موجود در آن سایت دسترسی خواهد داشت. ولی مشکل وقتی نمایان می‎شود که جوینده اطلاعات نشانی از سایت مورد نطر را نمی‎داند و یا اینکه برای دریافت اطلاعات خود سایت یا سایتهای خاصی را در نظر ندارد، بلکه خواستار همه سایت‎هایی است که بتوانند اطلاعات مورد نیاز وی را ارائه نمایند. اینجاست که جذابیت‎های جست و جو (search)، موتورهای جست و جو (search Engines)، ابرموتورها ی جست و جوMeta search Engines) ) ،‌عملگرها (operators) و مهارت‎های کاوش آشکار می‎شود و جست و جو در اینترنت به یک هنر و حتی یک علم یا فن تبدیل می‎شود(چنان که نوع نافرجام آن به عامل خستگی یا یأس تبدیل می‎شود)

مرکز امور رایانه‎ای دانشگاه امام صادق علیه السلام بر آن است از این شماره با ارائه تعاریف اولیه، معرفی ابزارهای جست و جو و تکنیک‎ها و استراتژی‎های کاوش،‌زمینه را برای بازیابی اطلاعات از اینترنت فراهم سازد.

تعاریف مقدماتی

1.          **طبقات اطلاعات:** داده یا data به یک عدد یا یک واقعیت که بیانگر مشاهدات خارج از متن بوده و به طور مستقیم ممکن است معنی دار نباشد اطلاق می‎گردد و با اطلاعات یا Information که از قرار گرفتن داده در یک بافت معنی‎دار حاصل می‎شود و اغلب به صورت یک پیام است، متفاوت می‎باشد. همچنین دانش یا knowledge چیزی است که ما آن را داریم و برای آن ارزش قایل هستیم که بر اساس پیام‎هایی که از طریق تجربه، ارتباطات واستثناها به دست آمده‎اند، تعریف می‎شود. ولی حکمت یا بینش یا Wisdom از کشف ارتباطات میان دانش‎های گوناگون حاصل می‎شود.

2.                    **جست و جو:** فرآیند بازیابی اطلاعات

3.          **بانک اطلاعاتی:** عبارت است از سیستمی رایانه‎ای(شامل نرم‎افزارها و سخت افزارها) جهت ذخیره و بازیابی اطلاعات با قابلیت جست و جو، ویرایش و اضافه کردن اطلاعات(با توجه به جزئیات پیش‎بینی شده). بانک‎ها می‎توانند اطلاعات را در فرمت‎هایی مختلف در خود نگهداری کنند(چکیده، نمایه،‌تمام متن،‌تمام تصویر،....)

4.                    **انواع جست و جو:** در اینترنت ممکن است در یک سایت و یا در یک صفحه انجام گیرد.

**1.ابزارهای جست و جو:** موتورهای جست و جو (search engines): به طور کلی موتورهای جست و جو نرم‎افزارهایی در محیط وب هستند که به طرق مختلف امکان جست و جو در اینترنت (یا بهتر است بگوییم در صفحات وب ) را فراهم می‎آورند. ذکر این نکته ضروری است که به دلیل حجم بسیار زیاد این صفحات،‌هیچ موتوری با همه نقاط قوت خود، قادر به جست و جو در همه منابع موجود در اینترنت نیست و هر موتور با توجه به توانایی‎ها و قابلیت‎های خود، تعدادی از این صفحات را می‎تواند جست و جو کند، با توجه به آمارهای غیر رسمی،‌حدود 10 میلیارد صفحه در وب جای دارد Northern light,Altavista, Hotbot و... هر یک به تنهایی تعدادی از این منابع وب را پوشش می‎دهند.

**امکانات موتورهای جست و جو**: گفته می‎شود که بیش از 2000 موتور جست و جو در وب طراحی شده است که هر یک از آنها،‌ویژگی‎های خاص خود را دارند و برخی از آنها به علّت دارا بودن توان بیشتر در پاسخگویی به نیازهای کاربران، بسیار پراستفاده‎اند.

به طور کلی امکانات طراحی شده در موتورهای جست و جوی مختلف به شرح زیر است:

- راهنمای کمکی

- گزینه‎هایی برای جست و جوی یک یا چند نوع از منابع اطلاعاتی

- جست و جوی کلید واژه‎ای

- جست و جوی عبارتی

- جست و جوی مفهومی

- پالایش جست و جو

- کوتاه سازی یا بریدن بخشی از واژه

- استفاده از عملگرهای بولی

- استفاده از علامتهای+ و- (اصلاح کننده‎ها)

- جست و جو در فیلدی خاص

- استفاده از گیومه

- محدود کردن جست و جو به مناطق جغرافیایی

- محدود کردن جست و جو به سایت‎های خاص

- ذخیره کردن راهبرد جست و جو

- بازبینی املای واژة مورد جست و جو

- جست و جو با حروف بزرگ

- استفاده از عملگرهایnear یا Adjacent

- نمایش نتایج جست و جو به صورت یکپارچه

- رتبه‎بندی درجه مرتبط بودن نتایج بازیابی شده

- ارایه گزارمان یا چکیده نتایج جست و جو

- حذف نتایج تکراری

تصمیم گیری در انتخاب موتور جست و جوی مناسب، برای اغلب کاربران بسیار دشوار است،‌به ویژه اینکه پایگاه داده‎ها و الگوریتم های جست و جو به کرات در حال تغییر هستند.

**2-سایت‎ها(پایگاه‎های) پرتال:** سایت‎های پرتال به بازدید کنندگان امکان می‎دهند تا تقریباً هر چیزی که در جست و جوی آن هستند در یک محل پیدا کنند. این سایت‎ها معمولاً حاوی اخبار، وقایع ورزشی، گزارش وضعیت هوا،‌و همچنین امکان جست و جو در وب هستند. پیشتر وقتی کاربران واژه پرتال را می‎شنیدند،‌اغلب به یاد موتورهای جست و جو می‎افتادند و در واقع موتورهای جست و جو، نوع خاصی از پرتال یعنی پرتال افقی هستند. این گونه پرتال‎ها، اطلاعات را در مجموعه بسیار گسترده از موضوعات مختلف، جمع‎آوری می‎کنند. نوع دیگر پرتال‎ها که به پرتال عمودی معروف است اطلاعات را در مجموعه خاصی از موضوعات،‌جمع‎آوری می‎کند. به عبارت ساده‎تر،‌پرتال‎های افقی، عمومی و پرتالهای عمودی، تخصصی هستند. خریدهای online وکلاً تجارت الکترونیک، بخش عمده‎ای از سایت‎های پرتال را به خود اختصاص می‎دهند. ضمناً پرتال‎های عمودی،‌اطلاعات بسیار زیادی را تنها دربارة موضوعی خاص در اختیار کاربر قرار می‎دهند. این گونه سایت‎ها، اغلب سایت‎های اجتماعی نیز نامیده می‎شوند. سایت‎های اجتماعی مملو از گروههای مختلف با علاقه‎مندی‎های ویژه هستند. بهترین سایت پرتال افقی،‌Yahoo می‎باشد.

**3-موتورهای جست و جوی نمایه‎ای:** شیوه کار بیشتر موتورهای جست و جو‌نمایه‎ای است؛ یعنی هنگام دریافت یک سایت یا یک صفحه، به صورت خودکار واژه‎های مندرج در آنها را بر اساس الگوریتم مشخصی، مثلاً بسامد هر واژه‎ و یا واژه‎های مندرج سر عنوان بخش‎های مختلف سایت، استخراج و در نمایه خود ذخیره می‎کنند.

عیب این شیوه در آن است که موتورهای جست و جو هر اندازه که هوشمند باشند نمی‎توانند مانند انسان به محتوای منابع پی ببرند و آن را در قالب یک یا چند واژه نشان دهند، لذا در بسیاری از موارد واژه‎های نامربوط که گویای محتوای سایت نیستند استخراج و به صورت کلید واژه ذخیره می‎شوند. دلیل بازیابی سایتهای نامربوط یا کم ربط در این گونه موتورهای جست و جو در همین مسأله نهفته است.

افزون بر آن، هنوز موتورهای جست و جو نتوانسته‎اند شکل نشان دادن جنبه‎های فرعی موضوعات و محتوای سایتها را حل کنند. بنابر این جست و جو در موتورهای جست و جوی نمایه‎ای می‎تواند بسیار مشکل و منوط به استفاده از کلید واژه مناسب و استفاده از عملگرها و ... ‎باشد.

**4-فهرست‎های راهنما‌و موتورهای جست و جوی موضوعی:** چون اطلاعات موجود در وب بسیار عظیم و فرّار است، داشتن فهرستی کاملاً جامع غیر ممکن می‎نماید،‌لذا تعدادی از مؤسسات و کاربران با گردآوری فهرست راهنمای وبDirectory) )ها، ابزار جست و جوگری را طراحی و قسمتی از این مشکل را حل نموده‎اند.

فهرست‎های راهنما منابع بسیار خوبی هستند،‌البته در صورتی که شیوة استفاده از آن را فرا گرفته باشید.

حتی اگر منابع اندکی را مرتبط با عنوان مورد نظر خود بیابید، این منابع باز هم شما را به سوی اطلاعات مورد نیاز خود راهنمایی خواهد کرد.

فهرست‎های رایگانی بر روی وب وجود دارند که دو راه برای استفاده از آنها وجود دارد.

بعضی از فهرست‎های راهنما، صفحات خانگی (Home page) دارند که در گروه‎هایی سازماندهی شده‎اند و برای استفاده از آنها باید صفحه خانگی را بارگذاری (Load) نمود و گروه دلخواه را انتخاب کرد. در هر گروه نیز، زیرگروه‎هایی وجود دارند که باید را آن قدر آنها را ادامه داد تا دقیقاً به آنچه مقصود است دسترسی پیدا کنم.

فهرست‎های راهنمایی وجود دارند که اجازه می‎دهند با استفاده از ماشین جست و جو (search Emgine) به دنبال مطلب خاص خود بگردید. استفاده از ماشین جست و جو بسیار ساده و آسان است. کافی است یک یا چند کلمه را در فرم مربوط تایپ و بر روی دکمه SUBMIT (یا Search) کلیک کرد و منتظر نتایج بعدی شد. اغلب ماشینهای جست و جو این اجازه را می‎دهند که از میان انواع گوناگون معیارهای جست و جو، کلید واژه مورد نظرتان را به روشی خاص فرمول بندی کنید. این گونه موتورها،‌یا به عبارتی راهنماها،‌سایت‎ها و صفحات را در ساختاری موضوعی مثلاً سلسله مراتبی یا درختی (subject-tree Structure) دسته‎بندی کرده و در زیر هر مقوله موضوعی بزرگ، گروه‎های موضوعی کوچکتر را آورده و به همین ترتیب،‌این دسته‎بندی تا رسیدن به موضوعات ریز و خاص ادامه یافته است. به عبارت دیگر، ساختاری سلسله مراتبی دارند که ارتباط هر موضوع را با موضوعات مادر(اعم) و موضوعات فرعی‎تر(اخص) نشان می‎دهند.

**5- ابرمورتورهای جست و جو (Meta Search Engines)** همان‎گونه که می‎توان حدس زد موتورهای جست و جو پیوسته روبه افزایش هستند. از سوی دیگر، کاربران حتی علی‎رغم رضایت از نتایج حاصل از جست و جو در یک موتور،‌تمایل به جست و جو در موتورهای جست و جوی دیگر را دارند، هر چند که این کار وقت‎گیر و هزینه‎بر است. برای رفع این مشکل از دهه 1990 به بعد، موتورهای جست و جویی طراحی شدند که قادر به جست و جوی همزمان در چندین موتور جست و جو هستند. ابرموتورهای جست و جو به خاطر اینکه خود پایگا‎ه داده‎ای در اختیار ندارند، کلید واژه‎ها یا عبارات وارد شده در کادر جست و جوی خود را همزمان به چندین پایگاه داده موتور جست و جوی منفرد و صفحات وب انتقال می‎دهند. سرعت بازیابی در این موتورها بسیار بالا است. زیرا این موتورها تنها به بازیابی مرتبط‎ترین مدارک از هر یک از موتورهای جست و جو قناعت می‎کنند. بسیاری از ابرموتورها، به مجرد رخ دادن تأخیر در دریافت پاسخ پرسش مطرح شده،‌ از موتور مورد نظر چشم می‎پوشند و ارتباط خود را با آن قطع می‎کنند. بدین ترتیب جامعیت جست و جو، قربانی سرعت در بازیابی می‎شود.

اصولاً متا(Meta) برابر واژه های «ورا و برتر» در فارسی و به این معناست که ابرموتور جست و جو (Meta search Engines) دارای امکانات و ویژگیهایی فراتر از امکانات یک موتور جست و جو می‎باشد.

واقعیت این است که تعدادی از موتورهای جست و جو که ادعای ابرموتور بودن دارند، حقیقتاً دارای ویژگی‎های خاص ابرموتور نیستند. در واقع برخی از آنها به خاطر این که قابلیت ارسال و نمایش ندارند، بیشتر شبیه مجموعه‎ای از موتورهای جست و جو یا راهنما و یا فهرستی بلند بالا از ابزارهای جست و جو هستند که برای هر یک از این ابزارها، یک کادر محاوره‎ای طراحی شده است. کاربران، پرسش خود را در کادر محاوره‎ای وارد می‎کنند. این پرسش به اولین موتور جست و جو ارسال می‎شود. پس از مشاهده نتایج، کاربر می‎تواند در صورت تمایل همان پرسش را به ترتیب به طور غیر همزمان در موتورهای جست و جو یا راهنماهای بعدی کپی یا درج کند.

ضمناً برای کار با موتورهای جست و جو آشنایی با دو مفهوم زیر لازم می‎باشد:

1.                   **Hit List :** پس از جست و جو، موتور جست و جو لیستی از اطلاعات مرتبط با عبارت جست و جو را نمایش می‎دهد که به آن Hit List می‎گویند.

2.                    **درصد مطابقت:** نتایج جست و جو در موتورهای جست و جو(مدارک یا صفحات بازیابی شده) بر اساس یک الگوریتم مشخص(که خاص هر کدام از موتورهای جست و جو است)، با توجه به ربط آنها به کلید واژه، مرتب و درجه‎بندی می‎شوند و برای آنها نمره‎ای (بر اساس حداکثر 100) تعلق می‎گیرد. این نمره در کنار آدرس منبع در hit list درج می‎شود.

**-----------**

**مبانی جستجودر اینترنت**

(قسمت دوم و پایانی)

**محمد رضا معبودیان اصفهانی**

در کنار انتخاب موتور جستجوی مناسب با توجه به نوع آن (نمایه ای یا موضوعی یا راهنما) که دارای بانک اطلاعاتی قوی و جامع درباره اطلاعات مورد نیاز باشد، داشتن یک یا چند کلید واژه و یا عبارت جستجوی مناسب برای کاوش در اینترنت الزامی است.

**مراحل تهیه عبارت جستجوی مناسب برای کاوش در اینترنت:**

**مرحله اول:** پیدا کردن اصطلاحات تخصصی، رایج و یا متعارف و ... دربارة موضوع مورد نظر. غالباً در هر رشته ای می توان از واژه نامه های تخصصی، فهرست عناوین تخصصی و یا سرعنوان های موضوعی و ... استفاده کرد.

**مرحله دوم:** تبدیل اصطلاحات تهیه شده در مرحله اول به عبارات جستجو. در این مرحله با استفاده از عملگردهای جستجو(Operators) ، بریده سازی (Truncatiion)، اصلاح کننده ها (Modifiers) و ... عبارت های جستجو تهیه و برای کاوش در موتورهای جستجو استفاده خواهد شد.

از آنجا که مرحله اول، تا حدی به میزان تجربه و تخصص و احاطه کاربر بر موضوع مرتبط است، در ادامه فقط به تشریح مرحله دوم (استفاده از عملگردها و بریده سازی و اصلاح کننده ها و ...) خواهیم پرداخت.

**عملگردهای جستجو:** از عملگرهای جستجو برای ارتباط دادن بین اصطلاحات و کلمات سازندة عبارت جستجو استفاده می شود. این ارتباط می تواند باعث گسترده شدن یا محدود شدن دامنه جستجو شود. (نحوه کار با عملگرهای جستجو با استفاده از جبر بولی میسر می باشد، لذا گاهی به عملگرهای جستجو، عملگرهای بولی نیز گفته می شود).

**عملگر AND** : هر گاه این عملگر بین دو کلمه یا اصطلاح در عبارت جستجو قرار گیرد، موتور جستجو، اطلاعات یا صفحاتی را بازیابی خواهد کرد که هر دو اصطلاح را شامل شود؛ مثلاً اگر عبارت جستجوی Internet AND Explorer (بدون توجه به ترتیب قرار گرفته کلمات) باشند. هرچه تعداد کلمات کلیدی که با AND به هم مرتبط می شوند، بیشتر باشد، نتایج کمتر، اما دقیق تری به دست خواهد آمد.

**عملگر** OR **:** این عملگر باعث بازیابی اطلاعات یا صفحاتی می شود که یک یا هر دو کلمة جستجو را شامل شود؛ مثلاً‌ اگر عبارت جستجوی Internet or Explorer را به کار ببرید، موتور جستجو، تمام صفحاتی را که حاوی یکی از این کلمات یا هر دوی آنهاست، تحویل می دهد. هر چه کلمات کلیدی بیشتری به وسیله عملگر OR ، در عبارت جستجو به کار برده شود، نتایج با دقت کمتر ولی حجم بیشتر ارائه خواهد شد.

عملگر NOT  : هر گاه این عملگر بین دو کلمه واقع شود اطلاعات یا صفحاتی بازیابی می شود که کلمه اول را در خود داشته باشد ولی حاوی کلمه دوم نباشد؛ مثلاً اگر عبارت Internet NOT Explorer را به کار ببرید، موتور جستجو، فقط صفحاتی را خواهد یافت که حاوی کلمه Internet و فاقد کلمه Explorer است. (در برخی از موتورهای جستجو به جای NOT  باید عبارت ANDNOT را به کار برد.)

علمگر ADJ : این عملگر اطلاعات یا صفحاتی را بازیابی می کند که هر دو کلمه را در کنار هم داشته باشد.

عملگر NEAR : برای جستجوی اطلاعات یا صفحاتی که در آنها دو کلمه  در یک جمله به کار رفته باشند از عملگر NEAR استفاده می شود.

نکته مهم: بعضی از موتورهای جستجو، ممکن است از بعضی عملگرهای جستجو پشتیبانی نکنند و یا روش‌های خاصی برای استفاده از آنها داشته باشند؛ از این رو بهتر است قبل از استفاده از عبارت های دارای عملگر جستجو، به راهنمای (help) موتور جستجو مراجعه نمایید.

نکته بسیار مهم: همان گونه که پیش تر اشاره شد، استفاده از عملگرهای AND و NOT و ADJ و NEAR باعث محدود شدن دامنه جستجو و بالطبع کم شدن نتایج و در عوض دقت بیشتر جستجو خواهد شد. همچنین استفاده از عمگر OR باعث گسترده شدن دامنه جستجو، زیاد شدن نتایج و دقت پایین خواهد شد. (البته گاهی اوقات لازم است از OR استفاده کنیم؛ مثلاً وقتی هیچ جوابی برای یک جستجو وجود ندارد, می توان  با OR محدوده جستجو را بازتر کرد و یا کلمات مترادف را با هم OR کرد.)

استفاده از پرانتز: در مواردی که از چندین عملگر برای جستجو استفاده شده و تقدم و تأخر عملگرها برای جستجو کننده اهمیت داشته باشد، می توان از پرانتز استفاده کرد. در این حالت جملات جستجویی که در داخل پرانتز قرار دارند، بر عبارات دیگر تقدم خواهند داشت.

بریده سازی (truncation) : با استفاده از این روش که معمولاً با درج ٭ در پایان کلمه انجام می شود کلیه مفاهیمی که از لحاظ املایی، ریشه یکسانی دارند، قابل جستجو هستند؛ مثلاً با جستجوی کلمه ٭ comput تمام اطلاعات حاوی computation, computer و ... بازیابی خواهد شد.

Wild card : این کاراکتر ( در اغلب موتورها کارکتر) می تواند جایگزین یک یا چند حرف در کلمة جستجو شده یا جایگزین هیچ حرفی نشود؛ مثلاً جستجوی cdonald؟m ، هر دو کلمه Mcdonald و Macdonald را بازیابی خواهد کرد.

اصلاح کننده ها (Modifiers) : اصلاح کننده (modifier) ، علامتی است که موتور جستجو را وادار می کند با کلمه ای که درست بعد از آن قرار دارد، رفتار خاصی داشته باشد. سه اصلاح کننده رایج در موتورهای جستجو به شرح زیر هستند:

1)     +: یعنی کلمه بعد از آن حتماً باید در صفحات وجود داشته باشد؛ مثلاً Iran + tehran + تمام صفحاتی را که دارای کلمات Iran و ‏Tehran (به هر ترتیبی) باشند، بازیابی خواهد کرد.

2)     -:یعنی کلمه بعد از آن در هیچ یک از صفحات نباید باشد.

3)     " ":عبارتی که داخل جفت کوتیشن قرار می گیرد، باید دقیقاً با همان شکل و ترتیب در صفحات موجود باشد؛ مثلاً‌ اگر عبارت “Internet Explorer" را جستجو کنید فقط صفحاتی  در فهرست نتایج ظاهر خواهند شد که هر دو کلمه Internet  و Explorer با همین ترتیب در آنها وجود داشته باشد.

**اسرار جستجوی موفق در اینترنت**

1-    اکثر جستجوگرها دو نوع محیط جستجو دارند: ساده و پیشرفته. در محیط جستجوی پیشرفته امکانات بیشتری وجود دارد که می توان با محدودتر کردن دامنه جستجو، ارتباط نتایج با موضوع جستجو را بیشتر کرد.

2-    سرعت به هنگام شدن (به روز رسانی) اطلاعات یک Search Engine در کاوش و نتایج آن بسیار موثر است.

3-    در عبارت‌های جستجو از s و es (و این گونه حروف گرامری ) استفاده نکنید.

4-    در عبارت جستجو، حرف اول اصطلاحات را بزرگ بنویسید؛ مثل Internet یا Explorer

5-    برای جستجوی نام یک شخص از «و» استفاده نکنید، زیرا اغلب موتورهای جستجو «و» را نادیده می گیرند؛ مثلاً از عبارت "Ali Hossini"جای  "Ali,Hossini"استفاده شود.

6-    در بانک‌های اطلاعاتی که با موتورهای جستجو در آن کاوش می شود, غالباً اطلاعات کلی، به صورت زیر دسته بندی می شود:

آدرس سایت، نام سایت، توضیح مختصر دربارة محتویات و فعالیت آن سایت و مجموعه ای از اطلاعات کلیدی که مرتبط با محتویات سایت است؛ بنابراین اگر کلیدواژه یا عبارت جستجوی شما، با یکی از این اطلاعات مطابقت داشته باشد آدرس آن صفحات یا سایت به همراه توضیح مختصر درخصوص محتویات و فعالیت آن سایت در موتور جستجو تهیه و به لیست موارد بازیابی شده اضافه می شود. (به این لیست hit list گفته می شود) دانستن دسته بندی های اطلاعات در بانک‌های اطلاعاتی، واژه گزینی در موتورهای جستجو را تسهیل می کند.

7-در روند کاوش در اینترنت، به هدف نخوردن جستجو یا گل نشدن آن به دو صورت است:

الف- هیچ اطلاعات یا صفحه ای پیدا نشود.

ب- بیشتر از 20 تا 100 صفحه پیدا شود.(ازدیاد صفحات، باعث سردرگمی کاربر خواهد شد.)

درصورتی که بیش از 20 تا 100 صفحه یا سایت در hit list  ظاهرشد توصیه می شود چند صفحه اول با توجه به توضیحات آن صفحات بازبینی شود. اگر آن چند صفحه با موضوع مورد نظر مرتبط نبودند، جستجو را دوباره با استفاده از عملگرها و اصلاح کننده ها و ... ادامه دهید.

اغلب با استفاده از عملگرها و اصلاح کننده ها (و دیگر تکنیک ها و ابزارها) می توان از کلید واژه نامناسب برای کاوش، یک عبارت جستجوی مناسب تهیه کرد.

8-اگر به املای کلمات در عبارت جستجو مشکوک هستید، می توانید از بریده سازی (استفاده از ٭ ) و یا Wildcard استفاده کنید.

9-در کنار موتورهای جستجوی اطلاعات, موتورهای جستجوی دیگری نیز در اینترنت وجود دارند؛ مثلاً موتورهای جستجوی اشخاص و موتورهای جستجوی برنامه های رایانه ای, موتورهای جستجوی موسیقی و فیلم و ...

10- برای استفاده از عملگرهای جستجو، حتماً AND ، OR ، NOT و ... را با حروف بزرگ بنویسید.

**معرفی موتورهای جستجو**

Altavista

آدرس موتور جستجو [WWW.altavista.com](http://www.altavista.com/)

آشنایی:

این سایت (پایگاه) یکی از بزرگترین و جامع ترین جستجوگرهای اینترنت است که اطلاعات بیش از 250 میلیون صفحه Web را در بانک اطلاعاتی خود جا داده است. اطلاعات این بانک اطلاعاتی دائماً به روزرسانی می شود و برای این به روزرسانی تمام صفحات html در آن به صورت «تمام متن» یا Fulltext نمایه می شود. در عین حال بزرگترین عیب این موتور ارائه نتایج بسیار به علت کثرت منابع می باشد.

**امکانات:**

○جستجو در 25 زبان

○جستجوی موضوعی (در گروه های هنر و سرگرمی Arts Entertainment ، اتومبیل «Autos » ، تجارت و امور مالی «Business Finance » ، کامپیوتر «Computer» ، بهداشت و پزشکی «Health Fitness» ، خانه داری و امور منزل «Home Family» ، اینترنت «Internet» ، اخبار و رسانه ها «News Media» ، سیر و سیاحت «Recreation Travel» ، مراجع «Reference» ، مناطق و کشورها «Regional» ، علوم «Science» ، فروشگاه‌ها و خرید «Shopping»، فرهنگ و جامعه «Society Culture» ، ورزش «Sports» و جهان «World») لازم است ذکر شود که این دسته بندی ها در اکثر موتورهای جستجوی عمومی رعایت می شود.

○حساسیت به حالت یا hello,HELLO,Hello)Case-Sensitive هر سه معادل هستند)

○ترجمه رایانه ای بین زبان‌های انگلیسی، فرانسه، ایتالیایی، پرتقالی، آلمانی، اسپانیایی

○کنترل بر نحوة نمایش نتایج

○امکان نوشتن جمله سؤالی در قسمت جستجو

○امکان جستجو در صفحات وب، اخبار، گروه‌های مباحثه، محصولات، تصاویر، فیلم و قطعات موسیقی

○مرتبه بندی نتایج جستجو

**کمبودها:**

○عدم امکان جستجوی مجدد از میان نتایج

○عدم اعلام درصد مطابقت عناوین پیدا شده با موضوع جستجو

**منطق:**

○عملگر OR به صورت پیش فرض میان کلیدواژه ها قرار دارد.

○امکان استفاده از عملگرهای NEAR,ANDNOT, AND نیز در حالت جستجوی پیشرفته (Advance) وجود دارد.

○بریدن کلمات خودکار نیست.

○استفاده از حروف بزرگ به معنای جستجوی دقیق برای تطابق کامل است. اما در صورت استفاده ازحروف کوچک، هر دو حالت بزرگ و کوچک جستجو می شوند.

○امکان استفاده از anchor در عبارت جستجو؛ با این فرمان صفحاتی یافت می شود که عبارت مورد نظر در آنها به صورت Hyper Link است.

مثال:anchor: Download IE5

○امکان استفاده از domain در عبارت جستجو؛ با این فرمان سایت هایی یافت می شود که domain آنها نوع مورد نظر باشد.

Domain یا قلمرو نشان دهنده نوع سایت و یا کشوری آن است مثل ir برای iran و uk برای انگلستان و ... ضمناً رایج ترین قلمروهای غیر کشوری عبارتند از Edu برای سایت‌های آموزشی / دانشگاهی ، Com  برای شرکت‌ها و مؤسسات تجاری، org برای سازمان‌ها و gov برای سایت‌های دولتی

مثال domain : edu

○امکان استفاده: از host در عبارت جستجو؛ با این فرمان صفحات یک رایانة خاص پیدا می شوند. مثال : host:www.isu.ac.ir

○امکان استفاده از: image در عبارت جستجو؛ این فرمان ازکاوشگر می خواهد صفحاتی را پیدا کند که دارای فایل تصویر مشخصی باشند:

مثال image; start

○امکان استفاده از:Link در عبارت جستجو؛ این فرمان که باید در جلوی آن نام یک سایت (URL) را وارد کرد، صفحاتی را پیدا می کند که با سایت به خصوصی پیوند دارند. مثال Link:www.zip2.com

○امکان استفاده از : text در عبارت جستجو: صفحاتی را پیدا می کند که شامل عبارت مورد نظر در هر قسمت صفحه باشد. مثال text:study

○امکان استفاده از: title در عبارت جستجو؛ صفحاتی را پیدا می کند که شامل عبارت مورد نظر در «عنوان» صفحه هستند. این عناوین در قسمت عنوان بیشتر مرور گرها ظاهر می شود. مثال title: sunset

○امکان استفاده از : URL در عبارت جستجو؛ با این زمان صفحاتی پیدا می شوند که عبارت یا کلمه مورد نظر در نشانی آنها وجود دارد.

مثال URL:isu

**فهرست نتایج جستجو:  این جستجوگر شامل عنوان، نشانی** URL**، دو خط اول توضیحات سایت، تاریخ آخرین به روزرسانی، حجم ( بر حسب بایت) و زبان اطلاعات است.**

Excite

**آدرس موتور جستجو:www.excite.ocm**

**آشنایی**

**بهترین مورد استفاده از این سایت،یافتن موضوعات روز و مسائلی است که در سطح وسیعی از آنها صحبت می شود. این جستجوگر درمیان 205 میلیون صفحه و ابزارهای رسانه ای و نیز اخبار** usenet **به جستجو می پردازد. این سایت همچنین مجموعه ای جامع و عمیق از موضوعات را از 150 هزار منبع در اختیار دارد. سایت** Excite **در زمینه جستجوی مفهومی بی نظیر است، زیرا این موتور با استفاده ترکیبی ازنمایه ها و موضوعات می تواند لغات کلیدی و موضوعات مختلف را برای شما جستجو کند. ضمناً** Exite **با استفاده از ارجاعات متون مرتبط به موضوع را هم می یابد.**

**امکانات**

**○جستجو در 11 زبان مختلف**

**○جستجو در میان فهرستی از کشورها یا** Domain **ها (مثل**org,com **. و ...)**

**○جستجوی مجدد از میان نتایج**

**○ارائه درصد مطابقت عنوان پیدا شده با عبارت جستجو**

**کمبودها**

**○حساسیت نداشتن به حروف بزرگ و کوچک**

**○امکان نوشتن جمله سؤالی در قسمت جستجو نیست**

**○مرتبه بندی نبودن نتایج جستجو**

**○نبود کنترل بر نحوه نمایش نتایج**

**منطق**

**○عملگر** OR **به صورت پیش فرض میان کلید واژه ها قراردارد.**

**○امکان استفاده از عملگرها** ANDNOT,AND **نیز وجود دارد.**

**در صورت وجود حروف بزرگ در ابتدای کلمات،** Excite **آنها را به عنوان نام شناسایی می کند.**

**○کلمات به صورت خودکار به عنوان پیشوند یا ریشه، در نظر گرفته می شوند.**

**○اصلاح کننده ها قابل استفاده هستند.**

**فهرست نتایج جستجو: شامل عنوان, نشانی** URL **, خلاصه‌ای ازمطالب و درصد ارتباط (مطابقت) با موضوع جستجو است.**

Google

**آدرس موتور جستجو** *[www.google.com](http://www.google.com/)*

**آشنایی:**

**این موتور از بزرگترین موتورهای جستجو وب در جهان است و گذشته از چندین سرمایه گذار خصوصی، دانشگاه استنفورد، شرکت** SUN **و شرکت های دیگری جزء سهامدارن آن هستند.** Google **با ادغام سخت افزارها و نرم افزارهای پیشرفته، همراه با الگوریتمی منحصر به فرد، قدرت پاسخگویی به سؤالات گوناگون کاربران را دارد. قلب نرم افزار آن** rankPage **نام دارد که در واقع فهرستی برای طبقه بندی صفحات وب است. نحوة کار** page rank **بدین صورت است که از ساختار بندی پیوند وسیع** ( vast link) **که در واقع ماهیت بسیار جالب و بی همتای وب است، به عنوان شاخصی از تعداد صفحات خصوصی استفاده می کند، مثلاً** Google **یک پیوند از صفحه** A **به صفحه** B **را همانند یک رأی از جانب** A **به** B **در نظر می گیرد و در نهایت به تعداد خالص این رأی ها یا همان پیوندهایی که از صفحه دریافت شده است، نگاه می اندازد. سپس صفحاتی را که بیشترین تعداد رأی از آنها صادر شده است، تحلیل می کند.**

**بدین ترتیب پایگاه های مهم اطلاعاتی، در هنگام طبقه بندی در نقطة بالاتری قرار می گیرند. همچنین** Google **تعداد ارجاعات کاربران به صفحات فوق به خاطر می سپارد و هر چه این ارجاعات بیشتر باشد، در دفعات بعدی، صفحات مورد نظر در نقطة بالاتری در طبقه بندی قرار می گیرند. در نهایت برای آنکه اطلاعات مورد نظر جستجو و بازیابی شود،** Google **از تکنیک «شمارش تعدد دفعاتی  که عبارت مورد نظر در یک صفحه دیده شده است» استفاده می کند.**

**امکانات:**

**○ جستجو به 50 زبان مختلف (جستجو در زبان عربی نیز امکان پذیر است؛ برای این کار در گزینه** goole in your language **زبان مورد نظر را انتخاب کنید.**

**○جستجوی عناوین ویژه** (topic specific search)

**○جستجوی موضوعی**

**○حساسیت به حالت یا** case sensitive

**منطق:**

**○عملگر** AND **به صورت پیش فرض میان کلید واژه ها قرار دارد.**

**○امکان استفاده از** OR **و** NOT **میسر است.**

**○امکان استفاده از اصلاح کننده ها و بریده سازی وجود دارد.**

**فهرست نتایج جستجو:** این **جستجوگر شامل عنوان نشانی** URL **, توضیحات و حجم (بر حسب بایت) است.**

**آدرس موتور جستجو [www.Yahoo.com](http://www.yahoo.com/)**

**آشنایی:**

Yahoo **بهترین موتور جستجوی موضوعی به شیوة سلسله مراتبی یا درختی است.** Yahoo **،پانصد هزار سایت, صفحه یا مدرک موجود در اینترنت را در یک ساختار سلسله مراتبی موضوعی و جامع (تقریباً 25 هزار دسته) آماده بازیابی کرده است. نتایج جستجو در** Yahoo **به صورت تقسیمات موضوعی** – **از موضوعات کلی تر به موضوعات جزئی تر** – **ارائه می شود. مزیت این شیوﺓ، بازیابی در آن است که جستجوگر با مقوله های موضوعی اعم و نیز اخص نسبت به موضوع مورد نظر آشنا می شود وحتی می تواند با انتخاب آنها، سایت ها و صفحات دیگری را بازیابی کند.** Yahoo **برای تسهیل بازیابی، امکان جستجوی کلیدواژه ای از موضوعات را نیز در اختیار می گذارد. ضمناً همراه با بازیابی سایت ها وصفحه ها، توصیفی کوتاه از آنها نیز ارائه می شود.**

**در عین حال، اشکالاتی به شرح زیر در** Yahoo **وجود دارد.**

**الف- در بسیاری از موارد، تغییر نشانی سایت ها از سوی پدیدآورندگان آنها در** Yahoo **اصلاح نشده (به روز نمی شود) و این امر موجب سردرگمی و اتلاف وقت کاربران می شود.**

**ب- حدس زدن اینکه مدارک یا سایت های مورد نظر کاربر، جزء کدام دسته یامقوله موضوعی است برای کاربر دشوار است. علاوه بر آن، برخی موضوعات در گروه یا دسته کاملاً مربوط ومناسبی قرار نگرفته است و این باعث سردرگمی بیشتر کاربر می شود. جستجوی کلیدواژه ای در** Yahoo **که برای کمک به کاربر در یافتن حوزه موضوعی مورد نظر استفاده می شود، در بیشتر موارد نتایج بسیار و بعضاً نامربوط را ارائه می کند. این سایت بزرگترین پرتال افقی در اینترنت است.**

**امکانات:**

**○جستجوی موضوعی (جستجوی سلسله مراتبی دارد.)**

**○نمایش نتایج در قالبها, دسته بندی ها و گروه های سایت های اینترنتی و صفحات امکان پذیر است.**

**○امکان جستجوی کلیدواژه‌ای در میان عناوین، گروه‌های خبری و آدرس های پست الکترونیک وجود دارد.**

**کمبودها:**

**○جستجوی حساس به حالت یا** Case-sensitive **ندارد.**

**○درصدد مطابقت هر عنوان باعبارت جستجو نمایش داده نمی شود.**

**○نبود کنترل بر نمایش نتایج**

**○عدم جستجوی مجدد از میان نتایج**

**○حساسیت به حروف بزرگ و کوچک در این موتور منظور نشده است.**

**○امکان نوشتن جمله سؤال در قسمت جستجو وجود ندارد.**

**منطق**

**○عملگر** AND **به صورت پیش فرض میان کلید واژه ها قرار دارد.**

**○امکان استفاده از عملگرهای** OR **و** NOT **موجود است.**

**○امکان استفاده از** t: **برای محدود کردن فهرست نتایج به فهرست عناوین** (title) **مثلاً :** t:Internet

**○امکان استفاده از** u: **برای محدود کردن فهرست نتایج به فهرست آدرس ها** (url) **مثلاً:** u:Internet

**امکان استفاده از اصلاح کننده ها وجود دارد.**

**فهرست نتایج جستجو :** در این موتور جستجو شامل عنوان، خلاصه، توضیحات سایت یا صفحه پیدا شده و آدرس سایت (url) است