

باسمه تعالی

جزوه آموزشی

# جستجوی پیشرفته در اینترنت

# ۱. مبانی جستجو در اینترنت

## ۱.۱. مقدمه

گاه جستجو برای یافتن سریع و صحیح اطلاعات در وب مانند این است که بخواهید یک سوزن را در انبار کاه بیابید، چرا که برای یک جستجوی ساده ممکن است تا چند صد میلیون مورد به شما ارایه شود. گاهی آنقدر با اطلاعات بی ربط در نتیجه جستجو مواجه می شویم که آنرا نیــمه کاره رها می کنیم.

اینترنت بیش از ۱۰ میلیون دامنه و ۹ میلیارد صفحه (تخمینی) را شامل می شود. تقریباً هر عبارتی را که در یک موتور جستجو وارد کنیم نتیجه هایی را در بر دارد. حتی اگر اشتباه تایپی داشته باشیم و خیلی طبیعی است اگر برای جستجوی یک عبارت چند میلیون نتیجه ظاهر شود. پرسش این است که: چگونه نتیجه ها را محدود به بهترین نتیجه ها کنیم؟ برای این کار باید روش درست این کار را بیاموزیم.

جستجو در شبکه اینترنت آسان، اما یافتن اطلاعات مفید و مناسب دشوار است. با وجود قابلیت هایی که موتورهای جستجو در امر بازیابی اطلاعات دارند انجام فرایند جستجو در آنها نیز نیازمند رعایت اصول و نکته هایی است و به مهارت های خاصی در امر جستجو نیاز است که بدون توجه به آنها انجام عمل جستجوی موفق تقریباً امکان ناپذیر و در بعضی موارد حتی ناممکن می باشد لذا در اینجا اصول و رهنمودهایی برای جستجوی موفق و مفید در موتورهای جستجو ارائه می گردد.

## ۱.۲. اصول و رهنمودهایی برای جستجوی موفق و مفید در موتورهای جستجو

- امروزه بیشتر ابزارهای جستجو در اینترنت برای معرفی امکانات و خدمات خود دارای راهنماهای کمکی می باشند که جستجوگر را راهنمایی می کنند. این راهنماها با واژه هایی چون ( Help ) یا ( search help ) در صفحه نخست موتور جستجو مشخص شده اند.
- برای دقیقتر کردن جستجو و همچنین محدود و یا گسترده کردن دامنه جستجوی خود از عملگرهای بولی ( AND, OR, NOT ) استفاده نمایند.
- در هنگام جستجو در موتور جستجو از کلمات کلیدی و درست استفاده کنید. به عبارت دیگر وقتی عبارت جستجوی خود را تهیه می کنید در آن عبارت یک یا چند کلمه کلیدی به کار ببرید. هر چه تعداد کلمات کلیدی در عبارت جستجو بیشتر باشد نتایج بهتری خواهید گرفت.
- در املای کلمه مورد جستجو دقت کنید و سعی نمایید که انواع مختلف املای واژه یا اصطلاح را جستجو نمایید. اگر املای دقیق کلمات را نمی دانید از کاراکترهای عمومی استفاده کنید. در بسیار از موتور های جستجو می توانید برای یافتن اطلاعات

در مورد کلماتی که از املای آنها مطمئن نیستند از کاراکترهای عمومی استفاده نمایید. در بیشتر موارد می توانید کاراکتر \* را به جای حروفی که در مورد آنها شک و تردید دارید قرار دهید.

- از مترادفهای مختلف واژه مورد نظر استفاده نمایید. به یاد داشته باشید که شما معمولاً در جستجوی یک واژه نیستند بلکه در جستجوی یک مفهوم و موضوع می باشید. پس برای رسیدن به نتایج بهتر در امر جستجو از مترادفهای واژه مورد جستجو استفاده کنید. برای مثال واژه های ( journals Electronic ) و ( Electronic magazine ) همگی مفهوم مجلات الکترونیکی را بیان می کنند و استفاده از هر کدام از آنها می تواند در امر جستجو به شما کمک نماید تا نتایج بهتری را بدست آورید.
- برای رسیدن به نتایج مطلوب و مفید از صفحه یا حالت جستجوی پیشرفته (Advanced search) استفاده کنید.
- از آنجا که هر یک از موتورهای جستجوی اینترنت دارای مجموعه ای از امکانات و قابلیت های کاوش خاص خود می باشند. برای رسیدن به نتایج مطلوب در امر جستجو بیش از یک موتور کاوش را جستجو کنید.
- برای انجام جستجو در موتورهای کاوش همچنین می توانید از قسمت ( Directory web ) که معمولاً در صفحه نخست موتورهای جستجو مشخص شده اند استفاده نمایید. در این قسمت رده های موضوعی مختلفی برای دسترسی به اطلاعات مشخص شده اند که در صورت کلیک بر روی هر کدام از آنها می توان به مقوله ها و موضوعات فرعی مرتبط با آنها دسترسی پیدا کرد. مثلاً در قسمت ( web Directory ) موتور جستجوی گوگل ( google ) تعداد ۱۶ مقوله و رده اصلی و پایه موضوعی نظیر Reference - computers and internet و غیره فهرست شده اند که جستجوگران می توانند با استفاده از آنها به مباحث و موضوعات مورد نظر خود دسترسی پیدا کنند.
- در عبارت های جستجو از s و es (و این گونه حروف گرامری ) استفاده نکنید.
- در عبارت جستجو، حرف اول اصطلاحات را بزرگ بنویسید؛ مثل Internet یا Explorer
- 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.
- 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...
- برای استفاده از عملگرهای جستجو، حتماً AND، OR، NOT و... را با حروف بزرگ بنویسید.
- برای این که ببینید چه سایت هایی به یک صفحه خاص لینک داده اند از عبارت زیر استفاده نمایید. البته تنها چند موتور جستجو چنین امکانی را می دهند. گوگل و MSN اینگونه هستند.
- link:www.par30theme.blogfa.com اگر عبارت: Link را قبل از آدرس صفحه مورد نظر وارد کنید موتور جستجو به شما نشان می دهد که چه کسانی به آن آدرس اینترنتی لینک (پیوند) داده اند. نتایج بازگشتی در این حالت هم شامل لینک های متنی و هم شامل تصاویر می شود.

### ۱.۳. مراحل تهیه عبارت جستجوی مناسب برای کاوش در اینترنت:

- **مرحله اول:** پیدا کردن اصطلاحات تخصصی، رایج و یا متعارف و... دربارهٔ 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 عملگرهای جستجو (Operators)، بریده سازی (Truncation)، 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 عملگرها و بریده سازی و اصلاح کننده ها و...) خواهیم پرداخت.
- عملگرهای جستجو:** از عملگرهای جستجو برای ارتباط دادن بین اصطلاحات و کلمات سازندهٔ عبارت جستجو استفاده می شود. این ارتباط می تواند باعث گسترده شدن یا محدود شدن دامنه جستجو 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. هر چه 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 کلمهٔ 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).
  - **عملگر 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:** 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 عبارت‌های دارای عملگر جستجو، به راهنمای (help) موتور جستجو مراجعه نمایید.

**نکته:** همان گونه که پیش‌تر 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 عملگر OR باعث گسترده شدن دامنه جستجو، زیاد شدن نتایج و دقت پایین خواهد شد.

**استفاده از پرانتز:** در مواردی که از چندین عملگر برای جستجو استفاده شده و تقدم و تأخر عملگرها برای جستجو کننده اهمیت داشته باشد، می‌توان 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 می‌شود کلیه مفاهیمی که از لحاظ املائی، ریشه یکسانی دارند، قابل جستجو هستند؛ مثلاً با جستجوی کلمه \*comput تمام اطلاعات حاوی computer.computation و... بازیابی خواهد شد.

- Wild card: این کاراکتر می‌تواند جایگزین یک یا چند حرف در کلمه جستجو شده یا جایگزین هیچ حرفی نشود؛ مثلاً جستجوی m?cdonald، هر دو کلمه McDonald و Macdonald را بازیابی خواهد کرد.

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

- +: یعنی کلمه بعد از آن حتماً باید در صفحات وجود داشته باشد. + مفهوم عملگر AND را داراست. شما می‌توانید چندین واژه را با عملگر + پیوند دهید و نتایج جستجو را هر چه دقیقتر نمایید. به ویژه در مورد برخی اطلاعات علمی استفاده از + بسیار ضروری است. بین کلمه اول و علامت + یک فاصله وجود دارد ولی بین + و دومین کلمه نه. برای نمونه با رشته روبرو اطلاعات مورد نظر خود را به دقت پیدا می‌کنید: «بهینه سازی +سایت +جستجو»

- - : یعنی کلمه بعد از آن در هیچ یک از صفحات نباید باشد. برای مثال با رشته زیر به دنبال مقاله‌ای در زمینه تحلیل سیستمها می‌گردیم که فاقد مطالب مهندسی برق باشد. زیرا بسیاری از مطالب تحلیل سیستمها در واقع زمینه مهندسی الکترونیک دارند ولی مطلوب ما نیستند. "مهندسی صنایع" + "تحلیل سیستمها" - "الکترونیک"

- " " : عبارتی که داخل جفت کوتیشن قرار می‌گیرد، 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.

## ۲. موتورهای جستجو

### ۲.۱. موتور جستجو چیست؟

به زبان ساده موتور جستجو (search engine) ابزاری است که به کمک آن می توان به جستجوی وب (اینترنت) پرداخت. امروزه با افزایش حجم اطلاعات بر روی شبکه اینترنت امر دستیابی به اطلاعات با اهمیت خاصی روبروست و بدون وجود چنین ابزارهایی عملاً امکان استفاده از بسیاری از منابع اینترنتی مقدور نخواهد بود.

### ۲.۲. بخش های مختلف موتورهای جستجو

- **عنکبوت:** عنکبوت با واریسی و پوشش صفحه های وب، پیوندهای موجود در هر صفحه به دیگر صفحات مربوط به آن صفحه را دنبال می کند. این روبات ها معمولاً هر چند وقت یک بار در اینترنت به جستجوی صفحات وب و ارتباط آن ها با صفحات دیگر می پردازند و در پایان، آنچه را پیدا کرده اند به نمایه می افزایند. گستردگی و عمق دسترسی به اطلاعات در هر موتور جستجو، بیش از هر چیز به ویژگی های نرم افزار خزنده آن بستگی دارد.
- **نمایه:** یک پایگاه اطلاعاتی است که اطلاعات نمایه سازی شده و مرتبط با سایت های وب در آنجا نگهداری می شود و قابل بازیابی است. ساختار نمایه، اندازه و حجم آن در موتورهای جستجو متفاوت است؛ به همین دلیل جستجو با کلیدواژه های یکسان در موتورهای گوناگون، نتایج نسبتاً متفاوتی را در پی خواهد داشت.
- **نرم افزار جستجو در نمایه :** برنامه ای است که در بین میلیون ها صفحه نمایه شده موجود در یک موتور جستجو، مطابق با پرسش جستجوگر و استراتژی های جستجو عمل می کند و اطلاعاتی را که با موضوع مرتبط باشد بازیابی می کند و نمایش می دهد.

### ۲.۳. انواع موتورهای جستجو

موتورهای جستجو سه دسته هستند:

- **موتورهای جستجوی عمومی (General Search Engines) :** این دسته از موتورهای جستجو به جستجو اطلاعات عمومی و در همه زمینه ها در اینترنت می پردازند .
- **موتورهای جستجوی تخصصی (Special Search Engines):** این دسته از موتورهای جستجو به جستجوی اطلاعات تخصصی و در یک یا چند زمینه خاص در اینترنت می پردازند.

- ابرموتورهای جستجو (Meta Search Engines): ابزارهایی هستند که به جستجوی اطلاعات در اینترنت و به طور همزمان در چند موتور جستجوی دیگر می پردازند.

در جدول زیر مثال هایی از انواع موتورهای جستجو بیان شده است.

موتورهای جستجوی عمومی	موتورهای جستجوی تخصصی	ابرموتورهای جستجو
Google	Ghatreh	MetaCrawler
Yahoo!	UTYP (Picture Search Engine)	<u>Vivisimo</u>
AltaVista	PhotoSearch	<u>Clusty</u>
AllTheWeb	Pipl	<u>Ixquick</u>
Gigablast	Iconfinder	<u>Dogpile</u>
HotBot		<u>Mamma</u>
Lycos		<u>Kartoo</u>

## ۲.۴. معرفی موتور جستجوی گوگل

این موتور از بزرگترین موتورهای جستجوی وب در جهان است. *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 *Page rank* نام دارد که در واقع فهرستی برای طبقه بندی صفحات وب است. نحوه 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 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 نقطه بالاتری قرار می گیرند. همچنین *Google* تعداد ارجاعات کاربران به صفحات فوق را به خاطر می سپارد و هر چه این ارجاعات بیشتر باشد، در دفعات بعدی، صفحات مورد نظر در نقطه بالاتری در طبقه بندی قرار می گیرند. در نهایت برای آنکه اطلاعات مورد نظر جستجو و بازیابی شود، *Google* از تکنیک «شمارش تعداد دفعاتی که عبارت مورد نظر در یک صفحه دیده شده است» استفاده می کند.

### امکانات:

- جستجو به ۵۰ زبان مختلف
- جستجوی عناوین ویژه

- جستجوی موضوعی

## 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.

## ۲.۵. معرفی موتور جستجوی یاهو

*Yahoo* بهترین موتور جستجوی موضوعی به شیوه 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 ۲۵ 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 شیوه، بازیابی در آن است که جستجوگر با مقوله های موضوعی اعم و نیز اخص نسبت به موضوع مورد نظر آشنا می شود و حتی 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 و صفحه ها، توصیفی کوتاه از آنها نیز ارائه می شود.

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

- 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.
- حدس زدن اینکه مدارک یا سایت های مورد نظر کاربر، جزء کدام دسته یا مقوله موضوعی است برای کاربر دشوار است. علاوه بر آن، برخی موضوعات در گروه یا دسته کاملاً مربوط و مناسبی 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 کلیدواژه ای در میان عناوین، گروه های خبری و آدرس های پست الکترونیک وجود دارد.



## کمبودها:

- درصد مطابقت هر عنوان با عبارت 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.

## ۲.۶ معرفی موتور جستجوی آلتاویستا

این سایت یکی از بزرگترین و جامع ترین جستجوگرهای اینترنت است که اطلاعات بیش از ۲۵۰ 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 ۲۵ زبان

- جستجوی موضوعی
- حساسیت به حالت یا Case-Sensitive.HELLO,hello (Hello هر سه معادل هستند)
- ترجمه رایانه ای بین زبان های انگلیسی، فرانسه، ایتالیایی، پرتغالی، آلمانی، اسپانیایی
- کنترل بر نحوه نمایش نتایج
- امکان نوشتن جمله سؤالی در قسمت جستجو
- امکان جستجو در صفحات وب، اخبار، گروه های 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:

- امکان استفاده از عملگرهای AND,ANDNOT,NEAR نیز در حالت جستجوی پیشرفته (Advance) وجود دارد.
- استفاده از حروف بزرگ به معنای جستجوی دقیق برای تطابق کامل است. اما در صورت استفاده از حروف 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 سایت های آموزشی / دانشگاهی، Com برای شرکت ها و مؤسسات تجاری، org برای سازمان ها و gov برای سایت های دولتی. مثال domain: edu
- امکان استفاده از host در عبارت جستجو؛ با این فرمان صفحات یک رایانه خاص پیدا می شوند. مثال: host:www.vums.ir

- امکان استفاده از image در عبارت جستجو؛ این فرمان از کاوشگر می خواهد صفحاتی را پیدا کند که دارای فایل تصویر مشخصی باشند: مثال image: start
- 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.

## ۳. راهنمای جستجو در موتور جستجوی گوگل

برای جستجو در گوگل باید کلیدواژه مورد نظر خود را در کادر جستجو وارد کنید و دکمه Enter را بزنید (یا بر روی دکمه Search کلیک کنید) تا صفحاتی که آن کلیدواژه در آن ها وجود دارد و یا صفحاتی که به آن موضوع مرتبط می شود بدست آیند. دقت داشته باشید که گوگل به طور خودکار از جستجوی کلمات بسیار رایج مثل a, for و how صرف نظر می کند. همچنین نوشتن کلمات با حروف بزرگ یا کوچک با یکدیگر فرقی نمی کنند.

### ۳.۱. جستجوی ترکیبی

اگر به دنبال صفحاتی هستید که دارای چند کلمه هستند جستجو به شیوه های زیر است:

- اگر دو یا چند کلمه را با یک کاراکتر فاصله از هم تایپ کنید، گوگل معمولاً صفحاتی را می آورد که ترجیحاً همه آن ها با هم در یک صفحه باشند و ترجیحات بعدی این است که این کلمات را با فراوانی بیشتر کلمات به فهرست نتایج وارد نماید.
- اگر به دنبال صفحاتی هستید که حتماً ۲ یا چند کلمه همزمان و با هم در یک جا باشند بین کلمات از + استفاده کنید. به این صورت که بین کلمات با یک کاراکتر فاصله از کلمه قبل از + استفاده کنید. مثلاً "اگر دنبال صفحاتی هستید که در آن ها هم music و هم classic باشد عبارت music + classic را در کادر جستجو وارد نمایید.
- اگر قصد جستجوی یک عبارت خاص را دارید باید این عبارت در "" قرار بگیرید. مثلاً اگر می خواهید عبارت classic music را جستجو کنید این عنوان به صورت "classic music" خواهد بود.

### ۳.۲. جستجوی پیشرفته

در صفحه اصلی گوگل، پیوند Advanced search مربوط به جستجوی پیشرفته می شود که مهمترین گزینه های آن به این شرح هستند.

- کادر **all these words** مربوط به وجود تمامی کلید واژه ها در صفحه نتایج است. یعنی کلماتی که در این کادر وارد می شوند همان نتایج وجود + در بین خود را در پی خواهند داشت.
- کادر **this exact wording or phrase** همان نقش " " را دارا می باشد. و کلیدواژه ها را به صورت یک عبارت در نظر می گیرد.
- کادر **one or more of these words** صفحاتی را جستجو می کند که حداقل یکی از کلیدواژه ها در آن باشد.
- کادر **any of these unwanted words** کلیدواژه هایی را که در آن تایپ می کنید از صفحات جستجو حذف می کند.

- قسمت **language** جستجو را به یک زبان خاص محدود می کند.
- در قسمت **File type** می توانید نتایج جستجو را به یک فرمت خاص از جمله **PDF** محدود کنید.
- اگر قصد دارید جستجو را فقط به یک سایت خاص محدود کنید، در بخش **Domain** آدرس این سایت را وارد کنید.

### ۳.۳ نکات جستجو در گوگل

#### نادیده گرفتن صفحات مشابه در نتایج جستجو:

توجه داشته باشید که گوگل در هنگام جستجو صفحات مشابه را حذف می کند و در لیست نمی آورد. این به این معنی است که اگر در دو سایت مختلف، صفحات نسبتاً یکسانی پیدا کرده باشد و یا دو صفحه ی مختلف از یک سایت را نسبتاً مشابه دیده باشد از آورد یکی از آنها خودداری خواهد کرد. اگر مایل بودید که صفحات مشابه را نیز ببینید، در آخرین صفحه نتایج جستجو در گوگل با متنی برخورد خواهید کرد. کافیست تا بر روی قسمت مشخص شده کلیک کنید تا جستجو دوباره انجام شود اما این بار با ذکر صفحات مشابه.

#### کلمات اصلی تر را بنویسید:

همانگونه که گفته شد برای بهتر نتیجه گرفتن از جستجو لازم است که کلمات اصلی را درنوار جستجو وارد کنید. برای اینکار هر قدر که می توانید، موضوع جستجو را ریزتر و معین تر نمایید. برای مثال بجای آن که در نوار جستجو تایپ کنید “نقاش” بهتر است که نام نقاش مورد نظرتان را تایپ کنید، مثلاً بنویسید “پیکاسو”. یا اینکه بجای “اتومبیل” بهتر است بنویسید “پژو” و یا ... همچنین دقت نمایید تا کلماتی را بکار برید که احتمال وجود توضیحات درباره آنها بیشتر باشد. مثلاً به جای نوشتن “گذراندن یک شب خوب در تهران” بهتر است تایپ کنید “هتل های تهران”.

#### از آوردن حروف اضافه خودداری کنید:

گوگل بطور پیش فرض فقط سایتهایی را لیست می نماید که تمام کلمات مورد جستجو در آنها باشند. و نیازی نیست که شما بین کلمات جستجو از کلمه پیوند دهنده “و” استفاده نمایید. به خاطر داشته باشید که وارد نمودن چنین کلماتی و کلماتی نظیر در، به، و ... بر نتیجه جستجو تاثیر خواهد گذشت. بطور مثال برای جستجو در مورد تعطیلات در شمال بجای تایپ نمودن “گذراندن تعطیلات در شمال” بهتر است تنها بنویسید: “تعطیلات شمال”.

خودِ گوگل هم برای بالا بردن کیفیت جستجوها بطور اتوماتیک کلمات معمولی و کاراکترهائی همچون “کجا” و “چطور” را از کلمات مورد جستجو حذف می نماید. چون این گونه کلمات باعث کندی عمل جستجو می شوند و تاثیری هم در بهبود جستجو ندارند و حتی کیفیت نتایج جستجو را کاهش هم می دهند.

## جستجوی تصویر:

جستجوی تصویر و عکس در گوگل و با جستجوی سایت فرق دارد، اگر دنبال عکس خاصی میگردید، بر روی گزینه image از بالای صفحه کلیک کنید و آنوقت در کادر مربوطه کلمه ی مورد نظر را تایپ کنید و کلید enter را بزنید

## جستجوهای طبقه بندی شده:

اگر مایلید که نتایج جستجوهای شما دقیق تر و علمی تر باشند بهتر است سری هم به [directory.google.com](http://directory.google.com) بزنید. در این آدرس فهرستی از سایتهای اینترنتی با توجه به موضوعات آنها موجود می باشد. مثلاً اگر دنبال سایتی در مورد سیاره زحل می گردید، می توانید با رفتن به شاخه Science و زیر شاخه Astronomy به زیر شاخه Saturn دست یابید. همچنین اگر در مورد خودروی Saturn جستجو می کنید می توانید سایتهای مرتبط با آنرا در شاخه Automotive بیابید.

## ۳.۴. عملگرها و متدهای جستجو در گوگل

اماعلاوه بر این تنظیمات، گوگل از یک سری عملگرها نیز برای محدود کرد جستجوها بهره می برد. شما بدون رجوع قسمت جستجوی پیشرفته و با دانستن این عملگرها می توانید عمل جستجو را هرچه سریعتر انجام دهید، در زیر به توضیح برخی از این عملگرها می پردازیم.

- **عملگر ~:** گاهی ممکن است که شما فقط به دنبال کلمه ای خاص نباشید و بخواهید که به دنبال کلمات مرتبط با آن نیز بگردید. برای این کار از این عملگر استفاده می شود. برای مثال اگر در کادر جستجو تایپ کنید «کیفیت غذا» در نتیجه سایتهایی ظاهر خواهند شد که کلمه «کیفیت» و کلمه «غذا» در آنها وجود داشته باشد. وحتی ممکن است که این دو کلمه اصلاً به هم ربط هم نداشته باشند. اما اگر تایپ کنید «~کیفیت غذا» دنبال سایتهایی می گردد که درباره کیفیت غذاها نوشته باشند.
- **عملگر SITE:** اگر می خواهید که عمل جستجو در سایت بخصوصی انجام گیرد، می توانید پس از وارد نمودن کلمه مورد نظر عملگر Site را آورده و پس از آن آدرس سایت را بیاورید. برای مثال برای جستجوی «پذیرش» در سایت دانشگاه استنفورد می توانید بنویسید [admission site:www.stanford.edu](http://admission.site:www.stanford.edu)
- **عملگر..:** برای اینکه جستجو را محدود به سایت هایی کنید که در آنها اعداد محدوده خاصی وجود داشته باشند، می توانید از این عملگر برای نشان دادن محدوده عددی مورد نظر بهره برید. مثلاً اگر به دنبال DVD Player های با قیمت ۲۵۰ تا ۳۵۰ دلار میگردید، بایستی اینگونه خواسته تان را بیان کنید (250..350 DVD player: توجه داشته باشید که بین دو نقطه عملگر هیچ فاصله ای نباشد).
- **متد filetype:** با استفاده از این متد می توانید به گوگل بگویید که در اینترنت به دنبال فایل هایی با پسوند تعیین شده بگردد. برای مثال می توانید این عبارت را در گوگل جست و جو کنید `filetype:txt` : و سپس گوگل لیستی از فایل های txt را در اختیار شما قرار می دهد. پسوند هایی که بیشتر به کار می آیند به شرح زیر است:

- پسوند doc و docx برای فایل های ورد
- پسوند ppt و pptx برای فایل های پاورپوینت
- پسوند PDF برای کتاب های الکترونیک
- پسوند dwg و dxf برای اتوکد
- پسوند max و ds و obj برای 3dsmax
- پسوند jpg و bmp و png و ... برای عکس
- **متد: link** با استفاده از این متد، گوگل صفحه هایی را که به یک آدرس مخصوص لینک داده اند جست و جو می کند . برای مثال [link:http://www.vums.ir](http://www.vums.ir) لیستی از سایت هایی که به سایت وارستگان لینک داده اند را به شما نشان می ده د.
- **متد: related** : با این متد، گوگل صفحه هایی را که به صفحه ی مورد نظر شما شبیه هستند برای شما لیست می کند . برای مثال با جستجوی عبارت [related:http://www.vums.ir](http://www.vums.ir) در گوگل، لیستی از صفحاتی که به home page سایت ذکر شده شباهت دارند در اختیار شما قرار می گیرد .

### ۳.۵. آشنایی با برخی از سرویس های گوگل

- **Google Scholar**: جستجوی علمی گوگل که برای جستجوی مقالات علمی طراحی شده است البته این سرویس هنوز در مرحله ی بتا می باشد ولی توصیه می شود که دانشجویان حتما از این سرویس استفاده کنند .
- **Google Sites**: با گوگل یک سایت واقعی بصورت کاملا رایگان بسازید و مدیریت آن را انجام بدهید و مطالب خود را در آن بگذارید.
- **Google Video**: جستجوی ویدئو از طریق سایت گوگل در سایتهای به اشتراک گذاری ویدئو و دانلود فیلم
- **Youtube**: معروفترین و بزرگترین سایت به اشتراک گذاری ویدئو مربوط به گوگل می باشد آمارهای مختلفی از میزان آپلود و تماشای ویدئوها در یوتیوب به گوش می رسد که یکی از دلایل استفاده از این سرویس است .

## ۴. معرفی موتورهای جستجو و پایگاه های اطلاعاتی تخصصی

### ۴.۱. معرفی برخی از موتورهای جستجوی اختصاصی

<p>اشخاص</p> <p><a href="http://www.anystudent.com">www.anystudent.com</a></p> <p><a href="http://www.bigfoot.com">www.bigfoot.com</a></p> <p><a href="http://whowhere.lycos.com">http://whowhere.lycos.com</a></p>	<p>پزشکی</p> <p><a href="http://www.achoo.com">www.achoo.com</a></p> <p><a href="http://www.cspinet.org">www.cspinet.org</a></p> <p><a href="http://www.derwent.com">www.derwent.com</a></p>	<p>هنر و ادبیات</p> <p><a href="http://adam.ac.uk">http://adam.ac.uk</a></p> <p><a href="http://www.ahip.getty.edu">http://www.ahip.getty.edu</a></p> <p><a href="http://vos.ucsb.edu">http://vos.ucsb.edu</a></p> <p><a href="http://www.recent.ir">www.recent.ir</a></p>
<p>علوم</p> <p><a href="http://www.derwent.com">www.derwent.com</a></p> <p><a href="http://www.knowledgeexpress.com">www.knowledgeexpress.com</a></p> <p><a href="http://www.redherring.com">www.redherring.com</a></p>	<p>صوت و تصویر</p> <p><a href="http://www.3dtermaker.com">www.3dtermaker.com</a></p> <p><a href="http://www.reel.com">www.reel.com</a></p>	<p>کاریابی</p> <p><a href="http://comfind.com">http://comfind.com</a></p> <p><a href="http://hotjobs.com">http://hotjobs.com</a></p> <p><a href="http://monster.com">http://monster.com</a></p>
<p>کامپیوتر</p> <p><a href="http://www.download.com">www.download.com</a></p> <p><a href="http://www.filez.com">www.filez.com</a></p> <p><a href="http://www.tucows.com">www.tucows.com</a></p>	<p>کودکان</p> <p><a href="http://www.disneyblast.com">www.disneyblast.com</a></p> <p><a href="http://www.planetzoom.com">www.planetzoom.com</a></p> <p><a href="http://www.yahooligans.com">www.yahooligans.com</a></p>	<p>کتاب ها و مقالات</p> <p><a href="http://www.amazon.com">www.amazon.com</a></p> <p><a href="http://www.bibliofind.com">www.bibliofind.com</a></p> <p><a href="http://www.borders.com">www.borders.com</a></p> <p><a href="http://www.ehow.com">www.ehow.com</a></p>
<p>ورزش</p> <p><a href="http://www.cnnsi.com">www.cnnsi.com</a></p> <p><a href="http://www.nba.com">www.nba.com</a></p> <p><a href="http://www.soccernet.com">www.soccernet.com</a></p>	<p>موسیقی</p> <p><a href="http://www.lyrics.ch">www.lyrics.ch</a></p> <p><a href="http://www.napster.com">www.napster.com</a></p> <p><a href="http://www.workdrecords.com">www.workdrecords.com</a></p>	<p>تجارت</p> <p><a href="http://www.bbb.org">www.bbb.org</a></p> <p><a href="http://www.bizweb.com">www.bizweb.com</a></p> <p><a href="http://www.infospace.com">www.infospace.com</a></p> <p><a href="http://www.eforosh.com">www.eforosh.com</a></p>



وب	اخبار	تعلیم و تربیت
www.devsearch.com	http://7am.com	http://cricir.syr.cdu
www.w3.org	www.refdesk.com	www.studyabroad.com
www.tile.net	www.newsindexs.com	www.meli.dist.maricopa.edu
مسافرت	آب و هوا	بازی و سرگرمی
www.cheaptickets.com	www.intellicast.com	www.etoys.com
<www.lowestfare.com	http://weathernoaa.gov	www.gamesages.com
www.travelocity.com	www.weather.com	www.toysrus.com

#### ۴.۲. معرفی برخی از موتورهای جستجوی عمومی

www.google.com	www.goto.com	www.ask.com	www.search.com
www.aol.com	www.infoseek.com	www.mamma.com	www.finderseeker.com
www.msn.com	www.excite.com	www.av.com	www.askgeeves.com
www.altavista.com	www.yahoo.com	www.lycos.com	www.mozdex.com
800go.com	www.webcrawler.com	www.hotbot.com	www.exalead.com
www.dogpile.com	www.webcrawler.com	www.metacrawler.com	www.comiran.com
www.scoot.co.uk	www.live.com	www.internet.com	www.looksmart.com
www.alltheweb.com	www.cyberpedia.com	www.webtaxi.com	www.searchuk.com

### ۴.۳. دانلود رایگان منابع اینترنت



این موتور جستجوی هوشمند فایل با ایندکس سازی اسناد موجود در اینترنت و با وارد کردن کلمات کلیدی می تواند فایل های صوتی، تصویری، اف تی پی، کتاب های الکترونیکی، لینک های رپیدشیر و لینک فایل های تورنت را در کوتاه ترین زمان ممکن در اختیار شما قرار دهد. از خصوصیات جالب این موتور جستجوگر برای فایل های صوتی و تصویری جستجو و پخش آنها به صورت آنلاین می باشد.

### ۴.۴. موتور جستجوی اشخاص



بدون شک موتورهای جستجو نقش مهمی در پیدا کردن مطالب مورد نظر ما ایفا می کنند، به عنوان مثال برای پیدا کردن مقاله یا مطلبی سری به گوگل می زنیم، اما همیشه نیاز های ما پیدا کردن مقاله نیست، احتمالا و شاید مطمئنا برای شما هم پیش آمده که به دنبال فردی حقیقی در دنیای نت بوده باشید. در اینجا نیاز به یک موتور جستجوی اشخاص احساس می شود. موتور جستجویی که با نوشتن نام و نام خانوادگی فرد مورد نظر اطلاعاتی به ما بدهد.

موتور جستجوی سی وی گجت برای ما چنین کاری انجام می دهد. با نوشتن مشخصات فرد و جستجو در این موتور مواردی همچون وب سایت های مرتبط، اسناد، تصاویر، ویدئو ها، بلاگ و نیوز از گوگل و همینطور اطلاعاتی راجع به حساب های کاربری در سایت فیس بوک و مای اسپیس و... را ارائه می دهد.

### ۴.۵. انتخاب موتور جستجو

جستجو در اینترنت می تواند به نتایج با ارزشی منتهی شود ولی لازمه یک جستجوی موفق استفاده از یک یا چند موتور جستجو و توجه به قابلیت های آن است زیرا اگر فقط از یک موتور جستجو برای هر موضوعی استفاده کنید نتایج خود را بسیار محدود کرده اید جدول زیر که موتورهای جستجو با توجه به کاربرد در آن لیست شده است، به شما کمک می کند با توجه به موضوع و نوع فایل و دیگر مسائل بهترین گزینه را انتخاب کنید.

با توجه به موضوع و نوع فایل	
AllTheWeb   AltaVista   Dogpile   Fazzle   FindSounds.com   FtpFind   Genie Knows   HotBot Advanced Search   Ithaki   Ixquick   Lycos Music Downloads   Mamma   Lycos Multimedia Search   MSN Search Advanced Search   Researchville: Audio   Singingfish	موزیک
AllTheWeb   AltaVista   The Amazing Picture Machine   Ditto   Dogpile   Fazzle   Google Image Search   HotBot Advanced Search   IceRocket   Ithaki   Ixquick   Mamma   MSN Search Advanced Search   Picsearch   Researchville: Images   Yahoo! News Image Gallery	عکس ها
All TheWeb   AltaVista   Dogpile   Fazzle   HotBot Advanced Search   Lycos Multimedia Search   Mamma   Researchville: Video	فایل های صوتی و تصویری
گزینه های جستجو	
Chubba   Clusty   Copernic   Don Pedro   FastSeeks   Fazzle   Ithaki   Ixquick   Mamma   MetaCrawler   ProFusion   Query Server   SurfWax   more...	جستجوی همزمان بوسیله چند جستجو گر در آن واحد
نتایج جستجو	
A9   Alexa Web Search   AltaVista   Google   Google Web Directory   Teoma اکثر موتورهای جستجو این خاصیت را در نظر می گیرند	نتایج بر اساس لینکها دسته بندی میشوند، مثلاً بر اساس تعداد لینکهایی که از آن سایت در صفحات وب وجود دارد نتایج جستجو را رتبه بندی میکند
Alexa Web Search   AOL Search   Ixquick   MSN Search   Teoma   ZapMeta	نتایج بر اساس شهرت سایت و رتبه آن در دایرکتوری ها و موتورهای جستجوی دیگر رتبه بندی می شوند

Accumo   AllThe Web   Clusty   Dogpile   Don Pedro   Exalead   iBoogie   Infonetware   metaEUREEKA   Kartoo   KillerInfo   Query Server   Mooter   Teoma   Turbo10   Vivisimo	نتایج بر اساس مفاهیم و نوع سایت رتبه بندی می شوند
AllTheWeb   AltaVista   Lycos	نتایج بر اساس تعداد بازدیدهایی که هر شخص از صفحات مختلف آن سایت انجام می دهد انجام می گیرد
جستجوهای اختصاصی	
2RSS.com   Blogdex   Blogdigger   Bloglines   BlogPulse   Blogwise   Bloogz   Feedster   Globe of Blogs   Popdex   Technorati	جستجوی بلاگ
FileIndexer.com   FileWatcher.org   Ftpsearchengines.com   Oth Net	جستجوی فایل در سایتهای ftp ) متن، فیلم، موزیک، نرم افزار و...
AlltheWeb: News   AltaVista - News   FindNews.org   Google News Alerts   InfoGrid   NewsNow   NewsTrove.com   Pandia Newsfinder   RocketNews   Topix.net   The WorldNews Network	اخبار (در آن واحد از چندین موتور جستجو استفاده می کند. )
Daypop   Memigo   NewsIsFree	اخبار
About   Academic Info   BUBL LINK   INFOMINE   Librarians' Index to the Internet   Resource Discovery Network   Scout Report Archives   Virtual Learning Resources Center   WWW Virtual Library   more...	دایرکتوری های حرفه ای

#### ۴.۶. معرفی برخی از پایگاه های اطلاعاتی و موتورهای جستجوی داخل کشور

www.irandoc.ac.ir (پایگاه اطلاع رسانی پژوهشگاه اطلاعات و مدارک علمی ایران)
www.noormags.com (پایگاه مجلات تخصصی علوم اسلامی و انسانی نور)
www.sid.ir (پایگاه اطلاعات علمی جهاد دانشگاهی)

www.magiran.com (بانک اطلاعات نشریات کشور)
www.nlai.ir (پایگاه اطلاع رسانی سازمان اسناد و کتابخانه ملی جمهوری اسلامی)
www.hamayesh.org (پایگاه خبری تحلیلی همایش ها، کنفرانس ها، سمینارها و نشست های ایران)
www.ketab.ir (خانه کتاب)
www.srlst.com (مرکز منطقه ای اطلاع رسانی علوم و فناوری)
www.articles.ir (بانک مقالات و مطالب علمی ایران به زبان فارسی)
www.maghaleh.net (بزرگترین بانک مقالات دانشجویی)
www.mydocument.ir (بانک مقالات فارسی)
www.iranhamayesh.com (پایگاه اطلاع رسانی همایش ها، نمایشگاه ها و جشنواره ها)
www.iausroc.ir (سایت دبیرخانه دائمی همایش ها)

## ۵. وب پنهان

### ۵.۱. وب پنهان چیست؟

علاوه بر مراجعه به موتورهای جستجوگر یکی از راههای دیگر جستجوی اطلاعات، استفاده از **وب پنهان** است. اما به راستی وب پنهان چیست؟

اطلاعات موجود در اینترنت را می توان به سه دسته زیر تقسیم بندی کرد:

- **اطلاعات رایگان و پیدا :** اطلاعات رایگان و پیدا اطلاعاتی هستند که در دسترس همگان قرار داده شده اند و با جستجو در موتورهای جستجوگر می توانیم آنها را بیابیم. موتور جستجوگر هر چقدر هم از پایگاه داده بزرگی برخوردار باشد، نمی تواند تمام اطلاعات وب را در خود داشته باشد.
  - **اطلاعات رایگان و ناپیدا :** اطلاعات رایگان و ناپیدا اطلاعاتی اند که هر چند دسترسی به آنان مستلزم هزینه ای نمی باشد، اما نمی توانیم از طریق موتورهای جستجوگر به آنها دسترسی داشته باشیم.
  - **اطلاعات تجاری:** اطلاعات تجاری اطلاعاتی هستند که برای استفاده از آن ها باید پولی معادل ارزش آن، پرداخت شود.
- در حقیقت بخش اعظم وب از دسترسی موتورهای جستجوگر دور است که به آن وب پنهان گفته می شود، و در بردارنده اطلاعات رایگان و ناپیداست. در مقابل می توانید وب نمایان را بخشی از وب بدانید که موتورهای جستجوگر می توانند به آن دسترسی داشته باشند و در نتایج جستجو به مراجعه کنندگان خود نمایش دهند.

### ۵.۲. بخش های مختلف وب پنهان

#### وب مات یا تاریک

بخشی از فضای وب نامرئی به وب مات موسوم گردیده که می توانسته مورد استفاده کاربران قرار گیرد، اما به دلایل زیر این اطلاعات در خارج از دسترس کاربران قرار گرفته و موتورهای کاوش نمی توانند آن ها را بازیابی کنند:

- از آنجا که اولاً محیط وب دائماً در تغییر است و هر روز منابع و اطلاعات جدید به آن افزوده می گردد و ثانیاً صفحاتی در وب وجود دارند که هیچ پیوندی بین آن ها با منابع دیگر برقرار نشده، خزنده های موتورهای جستجو قادر به یافتن این صفحات و همگام نمودن خود با این حجم عظیم اطلاعات نیستند.

- به دلیل محدودیت توانایی، نرم‌افزارهای خزنده فرصت کافی برای روزآمدسازی صفحات جدید وب را ندارند. موتورهای کاوش نیز امکان روزآمدسازی حجم عظیمی از اطلاعات و منابع جدید را ندارند و به همین دلیل بسیاری از این اطلاعات از حوزه موتورهای کاوش دور می‌مانند.
- محدودیت توان مالی بسیاری از موتورهای کاوش سبب گردیده که موتورهای کاوش نتوانند تمام صفحات وب سایت‌ها را نمایه‌سازی کنند، چرا که برای آن‌ها هزینه‌های زیادی دارد و بنابراین موتورهای کاوش بنا بر سیاست‌های خودشان، تنها بخشی از وب سایت‌ها یا لایه‌های بیرونی آن‌ها را نمایه‌سازی می‌کنند. بنابراین همیشه بخش عظیم لایه‌های درونی وب سایت‌ها پنهان می‌مانند.

## وب عمیق

بخش دیگری از وب نامرئی، به مجموعه‌ای از اطلاعات الکترونیکی اطلاق می‌شود که بسیاری از پایگاه‌های اطلاع‌رسانی، آن‌ها را از طریق شبکه جهانگستر وب در دسترس عموم قرار داده‌اند. برخی این اطلاعات را به رایگان، و برخی دیگر را با دریافت هزینه در دسترس عموم قرار می‌دهند.

مندرجات این پایگاه‌ها معمولاً خارج از حوزه جستجوی موتورهای کاوش قرار دارند هر یک از این پایگاه‌ها صفحه جستجوی مبتنی بر وب دارند. که امکان جستجو در آن‌ها برای کاربران را فراهم می‌کند، اما خزنده‌های موتورهای جستجو توان ورود به آن‌ها را ندارند و در نتیجه حجم انبوهی از اطلاعات، نمایه نشده باقی می‌ماند. بنابراین کاربر باید در این موارد از طریق موتورهای جستجو، پایگاه‌های مرتبط با موضوع خود را شناسایی کند و سپس، جداگانه به جستجو در آن‌ها بپردازد تا از دسترسی به وب عمیق باز نماند.

## وب خصوصی و وب ملکی

بخشی دیگر از وب نامرئی وجود دارد که چون اطلاعات موجود در آن جزو دارایی‌های شخصی یا خصوصی سازمان‌ها یا افراد می‌باشد، از حوزه دسترسی موتورهای جستجو پنهان است. مثلاً در برخی از سازمان‌ها و مؤسسات خصوصی یا دولتی، به دلایل امنیتی از اطلاعات مربوط به مسائل کاری و سازمانی و پرسنلی خود حفاظت می‌کنند اجازه دسترسی به آن‌ها را به دیگران نمی‌دهند و فقط کسانی که دارای اسم کاربر و گذرواژه هستند می‌توانند از آن‌ها استفاده کنند؛ این بخش، وب خصوصی محسوب می‌گردد

بخش دیگر، منابع اطلاعاتی از قبیل نشریات الکترونیکی مبتنی بر وب می‌باشند که دسترسی به آن‌ها از طریق پرداخت حق اشتراک و خرید محصولات اطلاعاتی شرکت‌های مختلف صورت می‌گیرد و «وب ملکی» نامیده می‌شود.

## اینترنت واقعاً پنهان

بخش دیگری از وب پنهان وجود دارد که بنا به مسائل فنی و ناکارآمدی ابزارهای جستجو، از دسترسی کاربران دورمانده است. بسیاری از موتورهای جستجو قادر به بازیابی اطلاعات متنی اچ‌تی‌ام‌ال هستند، ولی توانایی بازیابی فایل‌های پی‌دی‌اف را ندارند، یا

به دلیل کمبود منابع مالی و فنی از جستجوی فایل‌های غیرمتنی صرف‌نظر کرده‌اند. بنابراین منابع اطلاعاتی متنوعی نیز در وب وجود دارند که تنها به دلیل محدودیت‌های فناورانه یا مالی موتورهای جست‌وجو، از حوزه کاوش آن‌ها و در نتیجه از دسترس کاربران دور مانده‌اند

### ۵.۳. دلایل عدم بازیابی و نمایه‌سازی وب نامرئی توسط موتورهای جست‌جو:

- **دلایل فنی:** بسیاری از موتورهای کاوش به دلیل محدودیت‌های نرم‌افزاری توانایی روزآمدسازی اطلاعات جدید وب را ندارند. باید یادآور شد که هنوز هیچ موتور کاوشی ادعا نکرده است که قادر به گسترش حوزه کاوش خود به تمام محیط وب می‌باشد و همیشه این موتورها یک گام از سرعت روزافزون اطلاعات عقب‌تر هستند.
- **دلایل بودجه‌ای:** همانطور که قبلاً اشاره شد فرآیند نمایه‌سازی تمام صفحات وب، هزینه‌بر خواهد بود و موتورهای کاوش نیز بنا به محدودیت بودجه ناگزیرند فقط بخشی از وب سایت‌ها را نمایه‌سازی کنند.
- **دلایل حقوقی:** از آنجا که اطلاعات موجود در وب در دسترس عموم قرار می‌گیرد، بسیاری از افراد و سازمان‌ها به دلیل صرف بودجه‌های کلان در راه‌اندازی سایت‌ها و پایگاه‌های اطلاعاتی خود، حاضر نیستند این اطلاعات را به صورت رایگان در اختیار همه بگذارند.

### ۵.۴. واقعیاتی درباره وب پنهان

- بهترین موتورهای جست‌جو فقط قادر به در دسترس قرار دادن حدود ۱۶٪ اطلاعاتی هستند که بر روی شبکه جهانی وب قرار داده شده است، بنابراین ۸۴٪ اطلاعات خارج از دسترس موتورهای جست‌جو است.
- اندازه وب پنهان ۵۵۰ بلیون بزرگتر از وب آشکار ظاهر است.
- وب پنهان حاوی حدود ۵۵۰ بلیون مدرک است، در حالیکه در مقام مقایسه، با وب آشکار یک بلیون صفحه است.
- وب پنهان بزرگترین رشد مقوله اطلاعات جدید را در وب دارد.
- ۹۵٪ وب پنهان اطلاعات برای همگان قابل دسترسی است (اما از طریق موتورهای جست‌جو نمی‌توان به آنها رسید).
- بیش از نیمی از وب پنهان در پایگاه‌های اطلاعاتی معین است.

### ۵.۵. چرا موتورهای جست‌جو نمی‌توانند اطلاعات وب پنهان را بیابند؟

- محتوا در پایگاه اطلاعاتی ذخیره شده است
- پایگاه اطلاعاتی با پرداخت هزینه قابل دسترسی است



- محتوا در زمان واقعی ارائه می‌شود
- قالبهای محتوا
- ورود به سایتها، نیازمند اجازه ویژه است
- تعاملی بودن محتوا
- محتوا به عنوان نتیجه کاوش کاربر، بصورت پویا تولید می‌شود.
- سایتهایی که توسط پروتکل‌های بازدارنده روباتها مسدود شده‌اند
- اطلاعاتی که به تازگی به سایت افزوده شده و موتورهای جستجو هنوز به آنها دسترسی نیافته‌اند.

نمونه‌ها	توسط کاوشگرهای استاندارد یافت نشده است
پایگاههایی مانند: Library catalogs , PubMed.ERIC	محتوای پایگاههای اطلاعاتی (با درخواست کاوش کاربر، محتوا تولید می‌شود)
قالبهای پی.دی.اف، دیداری و شنیداری و هر قالب جدید دیگر	قالبهای محتوا
وضعیت آب و هو، قیمت سهام، وضعیت پرواز هواپیماها	اطلاعات بلادرنگ (زمان واقعی)
سایتهای: LexisNexis Academic, INFOTRAC.EBSOChost	دسترسی به خدمات نیازمند پرداخت هزینه است
همه سایتهایی که نیازمند عضویت هستند	دسترسی به سایت نیازمند رمز عبور یا ثبت نام است
سایتهای راهنمای مسافرت، سایتهای کاریابی	دسترسی به سایت نیازمند پر کردن فرم خاصی است

## ۵.۶ راههایی که دسترسی به وب پنهان را هموار می‌کند

در حال حاضر ابزارهایی به وجود آمده‌اند که منابع وب نامرئی را شناسایی، و کاربران را به سایتهای مناسب راهنمایی می‌کنند. این رویکرد توسط بزرگراه‌های اطلاعاتی و کتابخانه‌های مجازی پذیرفته شده است؛ بطوری که فقط توصیفی از پایگاه‌های اطلاعاتی و مجلات نامرئی را ارائه می‌کنند؛ مثل سایت Invisibleweb که فهرستی از منابع نامرئی را، و سایت Completeplanet که فهرستی از تقریباً ۴۰۰۰۰ پایگاه اطلاعاتی وب نامرئی را ارائه می‌دهند.

بهای استفاده از وب پنهان ابتدا باید آدرس یکی از آنها را به کمک موتورهای جستجوگر بیابید. به عنوان مثال اگر موضوع پزشکی مد نظر شماست، کافیست که در گوگل اینگونه جستجو کنید: پایگاه داده پزشکی یا پایگاه داده علوم پزشکی. در هر صورت کلمه "پایگاه داده" یا "Database" یک کلمه کلیدی است.

به هر حال اولین نکته در استفاده از وب پنهان این است که فراموش نکنید که "وب پنهانی" وجود دارد و صرف کمی وقت برای استفاده از آنها می تواند مکمل مناسبی باشد برای آنچه که از طریق موتورهای جستجوگری نظیر یاهو و گوگل می یابید.

#### راهنماها:

- Open Directory Project: [www.dmoz.org](http://www.dmoz.org): بزرگترین راهنمای وب که ویراستاری آن توسط افراد (غیر ماشینی) انجام می گیرد.

- Librarian's Internet Index: [www.lii.org](http://www.lii.org): پایگاه اطلاعاتی با حدود ۱۷۰۰۰ وب سایت با عناوین متنوع. با ارائه خبرنامه

- Infomine: <http://infomine.ucr.edu>: مجموعه پژوهشهای اینترنتی دانش پژوهان از دانشگاه کالیفرنیا جنوبی

- Digital Librarian: [www.digital-librarian.com](http://www.digital-librarian.com): سایت «کتابدار دیجیتالی» مارگارت اندرسون – کتابدار

#### کاوشگرها:

- BrightPlanet's [www.completeplanet.com](http://www.completeplanet.com)

- [www.kartoo.com](http://www.kartoo.com)

- [www.search-22.com](http://www.search-22.com)

- [www.freepint.com/gary/direct.htm](http://www.freepint.com/gary/direct.htm)

- <http://searchpdf.adobe.com>

- [www.closerlooksearch.com](http://www.closerlooksearch.com)

- <http://oaister.umd.umich.edu/o/oaister/>

#### پایگاههای اطلاعاتی خاص:

- [www.findarticles.com](http://www.findarticles.com) - a database of free articles

- [MagPortal.com](http://MagPortal.com) - a database of free articles

- [www.anywho.com](http://www.anywho.com) - "Finding People and Businesses".

- etc.. lawsuits.[www.findlaw.com](http://www.findlaw.com) - information on federal and state laws •
- U.S. Dept. of Education - cataloged .<http://www.thegateway.org> - Gateway to Educational Materials •
- and university web sites.. state.collection of educational materials from federal
- [www.nsd.gov](http://www.nsd.gov) - National Science Digital Library •
- [www.usa.gov](http://www.usa.gov) •
- [www.gpoaccess.gov](http://www.gpoaccess.gov) •
- <http://catalog.loc.gov> - Library of Congress •
- [www.ncbi.nlm.nih.gov/PubMed](http://www.ncbi.nlm.nih.gov/PubMed) - free Web version of Medline •
- [www.infousa.com](http://www.infousa.com) - company database of 12 million records •
- [www.singingfish.com](http://www.singingfish.com) - For audio and video searching •
- and websites . databases.<http://www.scienceresearch.com/> - Searches web-based science journals •
- not accessible through general purpose search engines
- [www.shadowtv.com](http://www.shadowtv.com) •

## ۶. فرهنگ اصطلاحات وب

- **DNS:** (سیستم نام حوزه) وظیفه ترجمه نام سرویس دهندگان را به آدرس IP دارد. فرض کنید آدرس سایت [www.startweb.ir](http://www.startweb.ir) را در کاوشگر تایپ می کنید. کاوشگر این نام را به سرویس دهنده DNS در ISP شما می فرستد. سرویس دهنده DNS با برقراری تماس با سرویس دهنده نام ریشه (root name server)، سرویس دهنده نام [ir](http://ir) را پیدا کرده و آدرس IP ی [startweb.ir](http://startweb.ir) را از سرویس دهنده [ir](http://ir) بدست می آورد. در مرحله بعد با DNS در [startweb.ir](http://startweb.ir) تماس می گیرد تا آدرس **IP** سرویس دهنده وب [www.startweb.ir](http://www.startweb.ir) را بدست بیاورد.
- **IP:** به هر کامپیوتر در اینترنت یک آدرس منحصر به فرد به نام IP (Internet protocol) اختصاص داده می شود. IP از چهار بخش تشکیل شده که با نقطه از هم جدا می شوند. (مانند ۱۲۵.۲۳۵.۵.۸۲) هر بخش می تواند از ۰ تا ۲۵۵ باشد.
- **FTP:** مخفف File Transfer Protocol، به معنای پروتکل انتقال فایل می باشد که یکی از سریعترین و در عین حال ساده ترین روش برای انتقال فایل از کامپیوتر بروی اینترنت یا همان هاست است. این سرویس تقریباً از طریق تمامی میزبانان وب در اختیار کاربران خود گذاشته می شود تا آنها بتوانند تا آنها بتوانند از طریق نرم افزارهای مخصوص این کار، برای انتقال فایل ها استفاده نمایند.
- **سیستم مدیریت محتوا یا CMS یا Content Management System:** مخفف عبارت Content Management System است که به معنای سیستم مدیریت محتوا می باشد و به نرم افزارهایی گفته می شود که نظام قابل مدیریتی را در ایجاد، ثبت، بروزرسانی و نظارت، تصویب و نشر داده ها (محتوا) بازیابی اطلاعات فراهم می آورند.
- **Domain Sub:** نامی است که قبل از Domain اصلی قرار می گیرد و با نقطه از آن جدا می شود. به عنوان مثال [team.startweb.ir](http://team.startweb.ir) یک Sub Domain است و [startweb.ir](http://startweb.ir) یک Domain است.
- **Domain:** برای داشتن سایت باید یک نام برای سایت انتخاب کنید تا از دنیای نت یک قسمتش را به نام خود کنید. و اسمی که برای سایت می نویسین همان دامین هست <<<====<<<< [www. Domain.com](http://www.Domain.com)
- **دانلود:** داندلود بر عکس آپلود می باشد. یعنی گرفتن فایل های مختلف از روی فضای مجازی بر روی کامپیوتر خود. که بنا بر حجم آن فایل و سرعت اینترنت ما می تواند مدت زمان متغییر داشته باشد.
- **آپلود:** آپلود به معنای قرار دادن فایل های ما اعم از عکس، موزیک، فیلم و برنامه ها و... از روی کامپیوتر و حافظه ها خود به اینترنت اعم از Attachment کردن برای یک میل تا قرار دادن در فضای سایت و یا فضاهای مجاری مجانی می باشد.
- **اسپم یا Spam:** کلمه spam در ایمیل به معنای ایمیل ناخواسته و فرستاده شده در حجم زیاد می باشد (unsolicited Bulk E-mail) ناخواسته به این معناست که گیرنده، هیچگونه تصدیقی برای دریافت این پیغام های ارسالی نداده بوده و فرستاده شده در حجم زیاد به این معنا که پیغام به عنوان قسمتی از مجموعه بزرگتری از پیغام ها که همگی در

بر گیرنده محتوی یکسانی هستند فرستاده شده است . یک پیغام زمانی اسپم ( spam ) تلقی می شود که هم ناخواسته و هم فرستاده شده در حجم و تعداد زیاد باشد . ایمیل ناخواسته یک امیل معمولی است ( بطور مثال اولین تماس با یک شرکت، آگهی های تماس، تقاضا های فروش، استخدام...) ایمیل فرستاده شده در حجم زیاد هم یک ایمیل معمولی است ( بطور مثال خبرنامه ها - تماس با مشتریان توسط یک سایت )