



## ابزارهای بازیابی اطلاعات در اینترنت

موسی یمین فیروز

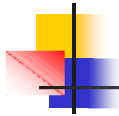
مدرس گروه آموزش عمومی دانشگاه علوم پزشکی بابل

yaminfirooz@gmail.com



## مقدمه

- اینترنت را به زباله دانی بزرگ تشبیه کرده‌اند که در آن قطعات الهاس یافت می شود منتهی اگر کسی روش زباله گردی در فضای اینترنت را نداند ، احتمالاً روزها و ماه ها در میان انبوه اطلاعات به درد نخور سرگردان خواهد شد و سرانجام بدون هیچ نتیجه مثبتی آن جا را ترک خواهد کرد
- معمولاً سخنی که از این دسته افراد شنیده می شود این است: **اطلاعات مناسبی وجود نداشت**. اما دیگردهای حرفه ای و کسانی که با روش ها و فن های جست و جو در اینترنت آشنایی دارند ، **هیچ وقت دست خالی از این فضای مجازی بر نمی گردند** آن ها همواره قطعات الهاس را در میان انبوه زباله های اطلاعاتی تشخیص می دهند .



## ابزارهای کاوش و جست و جو در وب

۱. فهرست راهنما یا دایرکتوری **Directory**
۲. موتور جست و جو **Search Engine**
۳. شاهراه اطلاعاتی **Gateway**
۴. ابر موتور جست و جو **Metasearch Engine**
۵. درگاه یا مدخل **Portal**
۶. وب نامرئی **Invisidle Wed**



## راهنماهاي موضوعي وب

- پایگاههای هستند که اطلاعات صفحات یا سایت های وب منتخب را توسط نیروی انسانی در پایگاه خود نهاده می کنند.
- **نکته:** در راهنماهای موضوعی وب فرآیند شناسایی، بررسی، تجزیه و تحلیل، ارزیابی و نهاده سازی صفحات توسط نیروی انسانی و بر اساس معیارهای از پیش تعیین شده انجام می شود.
- راهنماهای وب برای در دسترس قرار دادن اطلاعات صفحات یا سایت های وب اغلب از **نظام سلسله مراتب موضوعی** استفاده می کنند



## انواع راهنماهای موضوعی وب

- (۱) راهنماهای موضوعی عمومی: تقریباً کلیه موضوعات و حوزه های دانش بشری را پوشش می دهد. مانند:

Yahoo(<http://dir.yahoo.com>)

Looksmart([www.looksmart.com](http://www.looksmart.com))

Open Directory(<http://dmoz.org>)

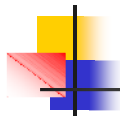


## انواع راهنماهای موضوعی وب

- (۲) راهنماهای موضوعی خاص: که منحصرآ برای دسترسی به اطلاعات حوزه ای خاص مانند فیزیک ، شیمی ، زیست شناسی ، کتابداری و ... ارائه می شوند. مانند:

Iranhoo([www.iranhoo.com](http://www.iranhoo.com))

یک راهنمای موضوعی به زبان فارسی مختص سایت های ایرانی است.



## نقش کتابداران در طراحی این راهنماها

- یکی از دلایل موفقیت راهنماهای موضوعی یاهو که هم اینک یکی از پرمراجعه ترین ابزار کاوش اینترنت به شمار می آید، استفاده از دانش چندین ساله علم کتابداری در سازماندهی سلسله مراتبی اطلاعات بوده است.
- متولیان اصلی یاهو در نخستین روزهای تاسیس آن **یک کتابدار با تجربه** به خدمت گرفتند.
- در حال حاضر بسیاری از راهنماهای موضوعی وب در حوزه های مختلف توسط کتابداران تهیه و گردآوری شده است که عبارتند از:



## راهنمای موضوعی Infomine

- که اطلاعات بیش از ۲۳ هزار سایت علمی را پوشش می دهد ، توسط کتابداران راه اندازی شده است. برای دسترسی به آن می توانید به آدرس زیر مراجعه کنید:

<http://infomine.ncr.edu>

یکی از مهمترین راهنماهای موضوعی که توسط کتابداران و متخصصان موضوعی در دسترس همگان قرار دارد Librarian index to Internet نام دارد:

[www.lii.org](http://www.lii.org)



## ۲) موتور های جستجو

■ پایگاههای اطلاعاتی قابل جستجویی هستند که از طریق برنامه های کامپیوتری به شناسایی و نمایه سازی خودکار صفحات وب می پردازند.

هر موتور کاوش داری سه جزء است:

۱) **روبات ها (Robots):** برنامه های خودکاری هستند که به طور پیوسته در فواصل زمانی معین (اغلب یک یا چند ماه) فضای اینترنتی را مبتنی بر ساختار فرایبوندی مورد جستجو قرار می دهند و به شناسایی و نمایه سازی صفحات وب جدید یا اصلاح اطلاعات می پردازند. در متون های مختلف از اصطلاحاتی چون **عنکبوت ها**، **خزنده ها** یا **کرم ها** نیز برای بیان این مفهوم استفاده شده است.



## اجزای موتور جستجو

۲) **پایگاه اطلاعاتی (Database):** فهرستی نظامند از صفحات وب است که اطلاعات مربوط به صفحات وب در آنجا ذخیره شده است. علاوه بر این نشانی دسترسی و توصیف مختصری از مدخلها را نیز ارائه می دهد.

به زبان دیگر هرچه توسط روبات جمع آوری شده است در پایگاه اطلاعاتی قرار می گیرد.

در متون های مختلف بجای این اصطلاح از عبارت **فهرست** یا **نمایه** نیز استفاده شده است





## اجزای موتور جستجو

۳) **نرم افزار بازیابی اطلاعات:** برنامه ای است که از طریق وارد کردن کلید واژه ها در فیلدهای مختلف آن ، می توان در میلیونها صفحه وب نمایه شده در پایگاه اطلاعاتی موتورهای کاوش به جستجو و بازیابی اطلاعات مورد نظر پرداخت.

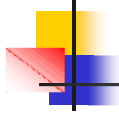
✓ منظور همان **رابط کاربر** و یا اولین صفحه ای که کاربر می تواند با پایگاه ارتباط برقرار کند. باید توجه داشت که این قابلیت در موتورهای جستجو متفاوت است.



## بنابر این فرآیند فعالیت موتورهای جستجو عبارتند از:

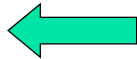
- (۱) شناسایی و نمایه سازی اطلاعات صفحات یا سایت های وب توسط روبات ها
- (۲) انتقال و درج اطلاعات گردآوری شده در پایگاه اطلاعاتی موتورهای کاوش
- (۳) جستجوی اطلاعات در پایگاه اطلاعاتی از طریق نرم افزار بازیابی اطلاعات
- (۴) بازیابی ، رتبه بندی و نمایش اطلاعات مبتنی بر اطلاعات موجود در پایگاه اطلاعاتی

## مزایای راهنماهای موضوعی نسبت به موتورهای کاوش



طبق بررسی های بعمل آمده ۴۱ درصد از کاربران اینترنت در آمریکا از راهنماهای موضوعی یا هو برای دسترسی به اطلاعات استفاده می کنند چرا؟

(۱) کیفیت بهتر اطلاعات: از آنجا که فرآیند نمایه سازی و تحلیل صفحات وب در راهنماهای موضوعی وب توسط نیروی انسانی انجام می شود و همین امر باعث شده است تا اطلاعات نمایه شده در پایگاه راهنمای موضوعی وب نسبت به موتورهای کاوش از کیفیت بالاتری برخوردار باشد.



## (۲) دسترسی بهتر به اطلاعات مرتبط



- رویکرد نمایه سازی انسانی در راهنماهای موضوعی و گنجاندن اطلاعات در قالب تقسیمات موضوعی باعث شده تا ریزش کاذب (بازیابی اطلاعات غیر مرتبط یا ناخواسته) در آنها نسبت به موتورهای کاوش بسیار کمتر باشد.
- ✓ همچنین در راهنماهای موضوعی خلاصه توصیف صفحات وب توسط متخصصان موضوعی نوشته می شود.

## ۳) صرف زمان کمتر برای دسترسی به اطلاعات

- ساختار موضوعی و سلسله مراتبی در راهنمای موضوعی وب به استفاده کننده امکان می دهد با انتخاب موضوعات اصلی و سپس موضوعات فرعی از وجود یا عدم وجود اطلاعات مورد نظر در راهنماهای موضوعی آگاه شوند. بنابراین با لیست بزرگی از نتایج روبرو نیستند.
- ۴) **سهولت مرور و بازیابی اطلاعات:** کاربر برای مرور مطالب فقط کافی روی موضوع مربوطه کلیک کند و نیازی به جستجوهای ساده و پیشرفته که نیاز به مهارت های خاصی دارد ندارد.

## محدودیت های راهنمای وب نسبت به موتورهای کاوش

- ۱) پوشش کم اطلاعات موجود در وب: چون نمایه سازی در این پایگاه بوسیله انسانها انجام می شود افراد قادر نیستند به ارزیابی محتوایی تمام صفحات وب بپردازند. (روزانه بیش از ۷ میلیون به صفحه وب افزوده می شود)
- ۲) روزآمد نبودن اطلاعات: نیروی انسانی توانایی بررسی و ارزیابی سایت های جدید را در زمان کوتاه ندارند. بنابراین قادر نیستند اطلاعات روزآمدی را در اختیار استفاده کنندگان قرار دهند.
- ۳) نیاز به آگاهی از ساختار سلسله مراتب موضوعی علوم: بعضی از کاربران از سلسله مراتب موضوعی علوم آگاهی ندارند و در بعضی از موارد راهنمای وب ها تقسیمات موضوعی متفاوت و غیر رایجی دارند.





## کدامیک بهتر است؟

- ✓ آنچه مسلم است نمی توان گفت کدامیک از ابزارهای کاوش (راهنمای موضوعی یا موتورهای جستجو) به طور مطلق بهتر از دیگری است. زیرا هر یک از آنها جایگاه و کاربرد خود را در فرآیند اطلاع یابی در اینترنت دارند و کاوشگران برای انجام جستجوی موفق باید منطبق با نیازهای خود و در زمان مناسب از هر یک از آنها بهره گیرند.
- ✓ ولی باید به این واقعیت توجه داشت که جامعیت و مانعیت به عنوان دو معیار ارزیابی نظامهای بازبازی اطلاعات است. و با توجه به مباحث مطرح شده در راهنماهای موضوعی جامعیت کم ولی مانعیت زیاد است ولی در موتورهای کاوش برعکس.



## نمونه هایی از موتورهای کاوش تخصصی

- موتورهای کاوش نشانی اشخاص: این ابزارها معمولاً امکان جستجوی نام اشخاص، نشانی پست الکترونیکی و نیز شماره تلفن آنها را در اینترنت در اختیار قرار می دهد. برخی از مهمترین آنها عبارتند از:

Bigfoot([www.bigfoot.com](http://www.bigfoot.com))

Internet address Finder([www.iaf.net](http://www.iaf.net))

Who Where([www.who哪里.lycos.com](http://www.who哪里.lycos.com))



## موتورهاي كاوش منابع صوتي و تصويري و چند رسانه اي



- اين موتورهاي كاوش عمدتا بمنظور جستجو و بازيابي فايلها و اطلاعات صوتي ، تصويري و چند رسانه اي مورد استفاده قرار مي گيرند. مهمترين آنها:

Scourhnet([www.scour.net](http://www.scour.net))

Imagesurfer(<http://isurf.interpix.com>)

Ditto([www.ditto.com](http://www.ditto.com))

## موتورهاي كاوش نرم افزارهاي موجود در اينترنت



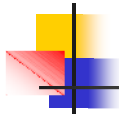
- موتورهاي كاوش زير امكان جستجوي سايتهاي را كه داراي سپاهه اي از نرم افزارهاي رايجان و قابل انتقال و يا فروشي هستند ، ارائه مي دهند:

[www.download.com](http://www.download.com)

[www.download.ir](http://www.download.ir)

[www.jumbo.com](http://www.jumbo.com)

[www.shareware.com](http://www.shareware.com)



## شاهراه اطلاعاتی Gate Way

- شاهراه اطلاعاتی در واقع نوع تخصصی دایرکتوری است. یعنی اطلاعات طبقه بندی شده در شاهراه محدود به یک موضوع می باشد.
- مانند شاهراه اطلاعاتی جامعه شناسی. اما دایرکتوری یا فهرست راهنما در همه زمینه ها به جمع آوری و طبقه بندی اطلاعات می پردازد.
- معمولاً شاهراه اطلاعاتی را استادان فن ارائه می کنند. از این نظر هرچند که ممکن است نشانی های موجود در آن اندک باشد، اما از اعتبار علمی بسیار بالایی برخوردار است.
- سایت [www.academicinfo.com](http://www.academicinfo.com) فهرستی از شاهراه های اطلاعاتی مربوط به رشته های مختلف علمی را دسته بندی کرده است.



## Invisible Web

- صفحات اینترنتی ای که ما به طور عادی می بینیم، وب سطحی نام دارند. افراد قبل از هر مرجع دیگری از وب برای جستجوی تحقیقاتشان استفاده می کنند. اما موتورهای جستجو لزوماً بهترین نتایج را نمایش نمی دهند. آنها اطلاعات مخفی یا عمیق موجود در اینترنت را نمایش نمی دهند.
- اصطلاح وب نامرئی یا وب "زُرف" به مخزن عظیمی از اطلاعات اطلاق می شود که موتورهای جستجو به آنها دسترسی مستقیم ندارند به طور مثال پایگاه داده کتابخانه های دانشگاهها، سایتهایی که برای مشاهده آنها نیاز به رمز عبور داریم و سایتهایی که به هر دلیل، موتورهای جستجو از فهرست بندی آنها، منع شده اند. بر خلاف صفحات قابل مشاهده (یعنی همین وبی که می توانید توسط موتورهای جستجو ببینید) اطلاعاتی در پایگاههای داده موجود است که اسپایدرها و رباتهای خزنده بدانها دسترسی ندارند.

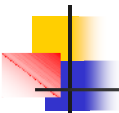


## وب نامرئی (ادامه)

### \* آمارهای جالب:

- اطلاعات موجود در وب ژرف ، بین ۴۰۰ تا ۵۵۰ برابر وب عمومی هستند که ما می بینیم
- وب ژرف حاوی ۷۵۰۰ ترابایت اطلاعات است ، در مقابل وب سطحی ۱۹ ترابایت ذخیره کرده است
- وب ژرف دارای نزدیک به ۵۵۰ بیلیون مستند می باشد در حالیکه این رقم برای وب سطحی ۱ بیلیون است
- بیش از ۲۰۰ هزار وب سایت ژرف وجود دارند

- جالب است بدانید که بیش از نیمی از محتوای وب عمیق در عنوان سایتهایشان از کلمه **Database** استفاده کرده اند و اگر این کلمه را در جستجوهایتان بکار برید ، امکان دسترسی تان را به این سایتهای افزایش می دهید.



■ پایان