



نمایه سازی

## فهرست مطالب

۵	دنیای وب و موتورهای جستجو.....
۵	درآمد زایی از تبلیغات با .....
۷	مقدمه ای بر نمایه سازی .....
۷	سه نکته مهم در ساخت موتورهای جستجو .....
۸	۱. موتور های جستجوی آزمایشی.....
۹	۲. موتورهای جستجوی تجاری .....
۹	انواع موتورهای جستجو .....
۹	موتورهای جستجوی مبتنی بر پیمایش وب و درون کاوی .....
۹	جستجوی مبتنی بر فهرست .....
۱۰	موتورهای جستجوی مختلط.....
۱۰	معماری کلی موتور های جستجو .....
۱۰	درون کاو یا خزنده (Crawler) .....
۱۰	کنترل درون کاو(Crawler Control) .....
۱۱	انباره صفحات (Page Repository) .....
۱۲	ماجول شاخص دهی (Indexer) .....
۱۳	Collection Analysis Module .....
۱۴	Utility Index .....
۱۴	موتور پرس و جو.....
۱۴	ماجول رتبه بندی .....
۱۴	درون کاوی صفحات وب .....
۱۵	انتخاب صفحات .....
۱۵	معیار مبتنی بر گرایشات کاربران .....
۱۵	معیار مبتنی بر شهرت صفحات .....
۱۶	معیار مبتنی بر محل قرار گرفتن صفحات .....

۱۶	مدل درون کاوی
۱۷	الگوی کاوش و توقف (Crawl & Stop)
۱۷	الگوی کاوش و توقف مبتنی بر مقدار آستانه (Crawl & Stop With Threshold)
۱۸	انباره ذخیره سازی
۱۸	گسترش پذیری یا Scalability تا بی نهایت
۱۸	پشتیبانی از دسترسی همزمان و دوگانه
۱۸	بهنگام سازی عظیم و توده ای
۱۸	صفحات منسوخ
۱۹	رتبه دهی و تحلیل لینک
۲۰	فایل robot.txt و کاربرد برچسب Alt tag
۲۰	روش های معنا بخشیدن به صفحات وب
۲۰	<b>XML</b>
۲۱	<b>URL</b>
۲۱	<b>URI</b>
۲۲	<b>RDF</b>
۲۳	نمایه
۲۳	دلایل نیاز به نمایه سازی
۲۳	اهداف نمایه سازی
۲۳	فراداده یا Metadata
۲۳	معرفی برخی متا تگ ها در طراحی وب
۲۴	زبان های نمایه سازی
۲۴	برخی از زبان های نمایه سازی
۲۵	انواع نمایه ها از لحاظ روش تنظیم
۲۶	انواع نمایه سازی از لحاظ روش تهیه
۲۶	دلایل نیاز به نمایه سازی خودکار
۲۶	آشنایی با برخی نمایه های قدیمی

۲۶	..... نمایه های گردان
۲۸	..... قنایط ضعف و قوت نمایه های گردان
۲۹	..... اصطلاح مرجح چیست؟
۲۹	..... ملاک انتخاب اصطلاح مرجح:
۲۹	..... ویژگی های یک نمایه خوب چیست؟
۲۹	..... نرم افزار های نمایه سازی
۳۰	..... <b>RoboHelp HTML Edition</b> آموزش نرم افزار
۳۳	..... لیست پروژه های انتخابی
۳۳	..... <b>RoboHelp HTML Edition</b> **لینک دانلود نرم افزار

## دنیای وب و موتورهای جستجو

بدون شک دنیای اینترنت بدون موتورهای جستجو معنایی ندارند. اقیانوسی از اطلاعات بدون داشتن ابزار لازم برای دسترسی به آنها فایده چندانی نخواهد داشت. اگر از روز اول ایجاد و شروع دنیای وب را تا به امروز بررسی کنیم می بینیم که نرخ تغییر و تحولات بسیار بالا بوده و روز به روز ابزار های جدیدتری در راستای کند و کاو بیشتر پدید آمده اند.



## درآمد زایی از تبلیغات با

شاید برای هر کاربر اینترنتی این سوال پیش بیاید که این همه خدمات فراوان و بطور غالب رایگان در دنیای اینترنت چگونه می شود که رایگان باشند؟ چرا شرکت بسیار بزرگ باید تا این اندازه خدمات متنوع و رایگان عرضه کنند؟

بدون آنکه به دنبال جواب باشیم به راحتی می توان تشخیص داد چه حجم ثروتی را این شرکت ها در این عرصه بدست می آورند. با اندکی مرور بر تاریخچه شرکتی همچون گوگل بیانگر این واقعیت می باشد. با وجود

اینکه شرکت گوگل در سال ۱۹۹۸ تاسیس و شروع به کار کرد اما در حال حاضر از همه شرکت های عرصه فناوری اطلاعات که سال ها و یا دهه ها قبل تر از گوگل شروع به کار کرده بودند پیشی گرفته است. مهمترین راه درآمد زایی برای این غول های فناوری اطلاعات کسب درآمد از طریق تبلیغات می باشد. به ظاهر که می نگریم میبینیم که شرکت گوگل پولی بابت خدماتی که ارائه می دهد دریافت نمی کند اما اگر در گوگل برای مثال واژه "میزبانی وب" را جستجو کنید مطابق شکل زیر یافت می شود که البته همیشه همین نتیجه نیست و ممکن است پویا باشد:

Google

میزبانی وب

All Images Videos Shopping Maps More Search tools

About 2,690,000 results (0.25 seconds)

هاست پر سرعت پویا سازان - یکی از ۱۰ هاست برتر کشور  
Ad [www.pouyasazan.org/](http://www.pouyasazan.org/) ▼  
تحویل آبی و خودکار هاست پس از پرداخت آنلاین

میزبانی وب - آریانیک  
[www.aryanic.com/services-webhosting.html](http://www.aryanic.com/services-webhosting.html) ▼ Translate this page  
میزبانی وب. شرکت آریانیک با در اختیار داشتن سرورهای متعدد با سیستم عامل های متفاوت و متنوع این امکان را در اختیار مشتریان خود قرار داده است تا مناسبترین سرور ...

هاست ایران | میزبانی وب - هاست لینوکس - خرید دامین - ثبت دامنه - سرور - طراحی  
[hostiran.net/](http://hostiran.net/) ▼ Translate this page  
هاست ایران بزرگترین مرجع هاستینگ و میزبانی وب ثبت دامنه و خرید دامین با قیمت عالی خدمات پشتیبانی و ماینیورینگ 24 ساعته.

میزبان پایتخت: هاست ، میزبانی وب ، دامین ، سرور مجازی  
[www.ehost.ir/](http://www.ehost.ir/) ▼ Translate this page  
میزبان پایتخت ارائه دهنده خدمات میزبانی وب، هاست و هاستینگ، میزبانی هاست، دامین، میزبانی نمایندگی، نمایندگی وب، سرور مجازی و سرور مجازی ابری می باشد.

این همان تبلیغاتی است که گوگل از شرکت های متقاضی دریافت کرده و برای کلید واژه تعیین شده "میزبانی وب" در این جستجو قرار گرفته مبلغی بابت آن از همان شرکت دریافت می کند.

از آنجا که این جستجو ها در تمام دنیا به زبان ها، گرایش ها، زمینه ها و موضوعات مختلفی بوده لذا گستره بسیار وسیعی از کلید واژه ها و کاربران را در بر می گیرد. شاید امکان تخمین دقیق این موارد امکان پذیر نباشد چرا که موتور های جستجو این آمار و اطلاعات را بیرون نمی دهند اما می توان به این حقیقت پی برد که چه درآمد میلیارد دلاری از همین طریق بطور سالانه نسبب آنها می شود. حال با این اوصاف باید ببینیم که موتورهای

جستجو در دنیای اینترنت چگونه جستجو می کنند و اطلاعات در این پهنه چطور ذخیره می شوند! همچنین نباید فراموش کرد که ماهیت پویا و تغییر پذیری موتورهای جستجو نسبت به تغییرات و ارتقا نحوه ذخیره، طبقه بندی و هویت اسناد اینترنتی سبب گردیده که سلطه بی چون و چرایی بر اینترنت داشته باشند.

## مقدمه ای بر نمایه سازی

تصور کنید وارد یک کتابخانه با هزاران و یا میلیون ها جلد کتاب می شوید. دریایی از دانش در پیش روی شماست اما دسترسی به آنچه که مورد نظرتان می باشد آسان نیست. در این لحظه شما نیازمند یک موتور جستجو هستید که با بررسی موضوعات مورد نظر شما هرآنچه را که تمایل دارید برای شما پیدا کند. موتورهای جستجو می توانند متفاوت باشند از اینکه گزینه مورد نظر شما در یک کتابخانه است یا در یک فروشگاه بسیار بزرگ و یا ... .

## سه نکته مهم در ساخت موتورهای جستجو

### ۱. خزیدن عمیق

موتور جستجو در لینک های موجود بر روی سایت های اینترنتی می خزد و داده های هر صفحه را به دیتابیس خود اضافه می کند. هرچه یک موتور جستجو لینک های بیشتری را پیمایش کند و این کار را سریع تر انجام دهد، داده های بیشتری در زمان کمتری به دست خواهد آورد.

### ۲. نمایه سازی

پس از جمع آوری داده ها، موتور جستجو باید آن ها را نمایه سازی کند. بدین معنا که هرچه با معیارهای نمایه سازی مطابقت بیشتری داشته باشد جستجوی کاربران نتایج مرتبط تر و موثرتری را در پی خواهد داشت.

### ۳. رتبه دهی

رتبه دهی یکی از ابتکارات شرکت گوگل بوده است. آنها به هر صفحه با توجه به معیارهایی که در اصلاح SEO گفته می شود رتبه بین 0 تا 10 می دهد.



موتورهای جستجو از لحاظ پشتیبانی های مالی و نیروی انسانی به دو دسته تقسیم می شوند:

### ۱. موتور های جستجوی آزمایشی

این دسته از موتورهای جستجو در دانشگاه ها یا موسسات پژوهشی شکل می گیرند و عموماً با هدف آزمایش یک روش جدید، ارزیابی نوع عملکرد و برآورد کیفیت نتایج یک الگوریتم پیاده سازی می شوند ولی از آنجاکه ایجاد یک سیستم بزرگ که بتواند گستره ی کل اینترنت را در پوشش بگیرد به منابع انسانی و مالی بسیار وسیع و مدیریت کلان نیاز دارد، این دانشگاه ها و موسسات نمی توانند رسالت اصلی خود را رها کرده و به بالنده کردن پروژه هایی پردازند. به همین دلیل از میان این طرح های آزمایشی هرکدام که سربلند و پیروز بیرون بیایند، از پوسته دانشگاهی خود بیرون آمده و وارد عرصه تجاری می شوند.



## ۲. موتورهای جستجوی تجاری

این رده از موتور های جستجو در پهنه کل اینترنت به جستجو و راهنمایی کاربران می پردازند و منابع مالی مورد نیاز برای ارتقاء و جلب رضایت کاربران را از طریق درآمد های تبلیغاتی و پشتیبانی مالی افراد حقیقی و حقوقی، تامین می کنند. روشن است که موتور های جستجو برحسب رضایت کاربران، بیننده های بیشتری را به خود جلب می کنند و جذب مخاطب و ترافیک بیشتر یک وبسایت مساوی است با افزایش بهای تبلیغات، سودآوری بیشتر و بالارفتن ارزش سهام شرکت صاحب آن. بدین ترتیب حلقه بسته "رضایت مخاطبین، افزایش سود، تامین منابع مالی و نیروی انسانی بیشتر، ارتقا خدمات و رضایت بیشتر مخاطبین" شکل می گیرد و در این حلقه کیفیت خدمات موتورهای جستجوی تجاری بهتر و بیشتر می شود. بزرگترین مشکلات ساخت موتورهای جستجو را می توان از یک طرف "ناهمگونی و حشمتناک اسناد و محتویات جای داده شده در وب" و از طرف دیگر "راضی نگه داشتن انبوهی از کاربران بی تجربه و پر توقع" خلاصه کرد.

### انواع موتورهای جستجو

#### موتورهای جستجوی مبتنی بر پیمایش وب و درون کاوی

بطور خودکار و هوشمند کل وب را پیمایش و مجموعه اسناد و پرونده های موجود را دریافت و رده بندی می کنند. جستجوی موارد مورد نظر کاربران بر اساس شاخص های تهیه شده صورت می گیرد. فرآیند دریافت، ذخیره، رده بندی و شاخص دهی در موتورهای جستجوی مبتنی بر پیمایش و درون کاوی بر اساس الگوریتم های هوشمند بصورت خودکار انجام می گیرد.

#### جستجوی مبتنی بر فهرست

با دخالت مستقیم و نظارت صاحبین اسناد و مستندات وب، ثبت، سازماندهی و اداره می شوند. بعنوان مثال در یک موتور جستجوی فرضی، کل فضای اسناد و اطلاعات موجود جهان در چند ده سطح کلان تقسیم بندی و تفکیک می شود. مانند رده های ورزش، هنر، تفریح، اخبار، هواشناسی، بورس و ... یا صدها رده ی دیگر تعریف می شود.

## موتورهای جستجوی مختلط

تلفیقی از یک سیستم فهرست غنی و یک موتور جستجو هستند تا دقیق ترین و کاملترین پاسخ را برای جستجوگران فراهم آورند. با وجود آنکه در سیستم فهرست، سازماندهی دقیق و نظم حاکم است ولی باز هم کاربران آماتور را راضی نمی کند چرا که علاقه ای به مسیر های تو در تو ندارند. تلفیق یک یا چند سیستم فهرست با یک موتور جستجوی قدرتمند که بتواند هوشمندانه این فهرست ها را زیر و رو کند و کاربران را به مقصود خود برساند بسیار مطلوب است.

## معماری کلی موتور های جستجو

موتور های جستجو متعدد و استراتژی های آنان متفاوت است و برخی از آنها به دلیل تجاری و امنیتی سعی می کنند الگوریتم های درونی شان را پنهان نگاه دارند. ابتدا اجزای این معماری را معرفی میکنیم و سپس در بخش های جداگانه جزئیات عملیاتی آنها را بررسی خواهیم کرد:

## درون کاو یا خزنده (Crawler)

تمام موتورهای جستجو به ماجولی عملیاتی به نام درون کاو متکی هستند که پهنه وسیع وب را بطور زمان بندی شده پیمایش می کنند و به جمع آوری صفحات بی شمار ذخیره شده در گوشه و کنار وب می پردازند. درون کاو از تعداد قابل توجهی برنامه کوچک تشکیل شده که هر کدام با هسته یک وب سایت خاص آغاز می کنند و پس از دریافت آن صفحه و تحویل به انبار صفحات بر اساس لینک های درون صفحه جاری، صفحات پیوند خورده را نیز به همین روال استخراج و ذخیره می کنند. این برنامه های کوچک، درون کاوی خود را تحت کنترل ماجول کنترل درون کاو آغاز و بر اساس استراتژی این ماجول کنترلی به پایان می رسانند.

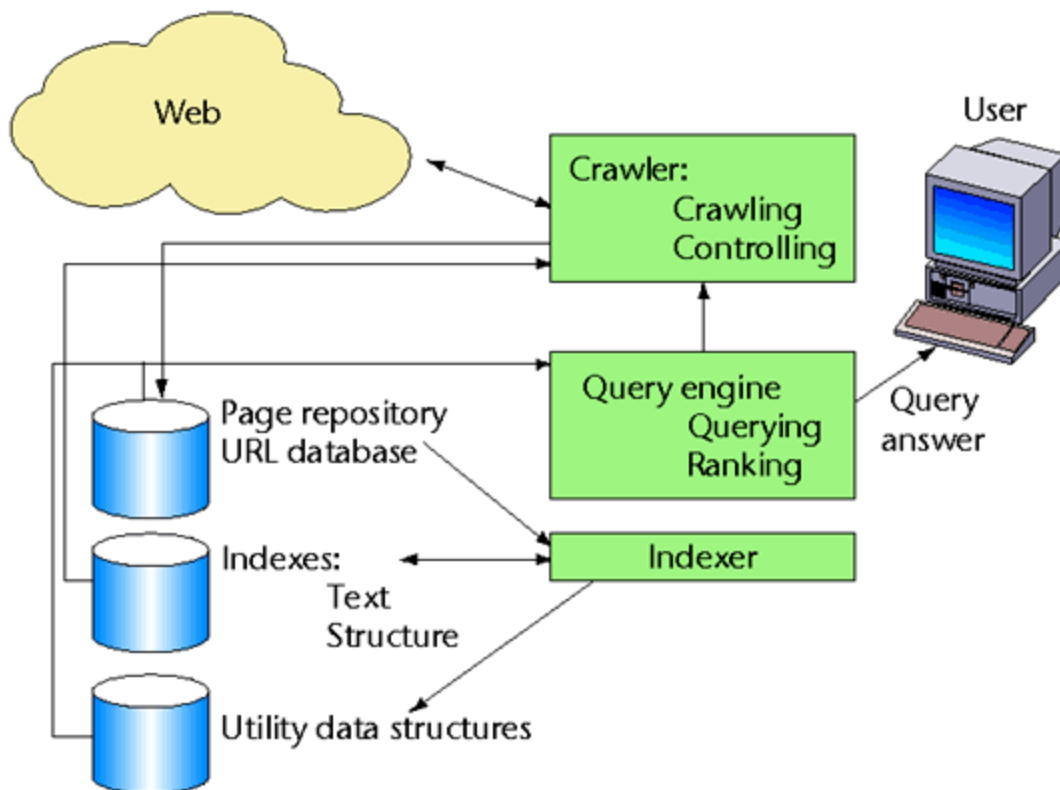
## کنترل درون کاو (Crawler Control)

این ماجول موظف است که مجموعه برنامه های درون کاوی را با تحویل URL مورد نظر، راه اندازی کرده و به کار درون کاوی بگمارد. ماجول کنترل درون کاو تعیین می کند که پس از دریافت صفحه جاری کدامین صفحه باید ملاقات و دریافت شود. چرا که یک صفحه واحد به صفحات متعددی لینک دارد. تزریق لینک هایی که باید به ترتیب ملاقات شوند بر عهده کنترل درون کاو است و در طرف مقابل ماجول خزنده موظف است تمام

آدرس ها یا به عبارتی URL های موجود را از درون صفحات دریافتی استخراج و آنها را برای تصمیم گیری در اختیار ماجول کنترل درون کاو بگذارد.

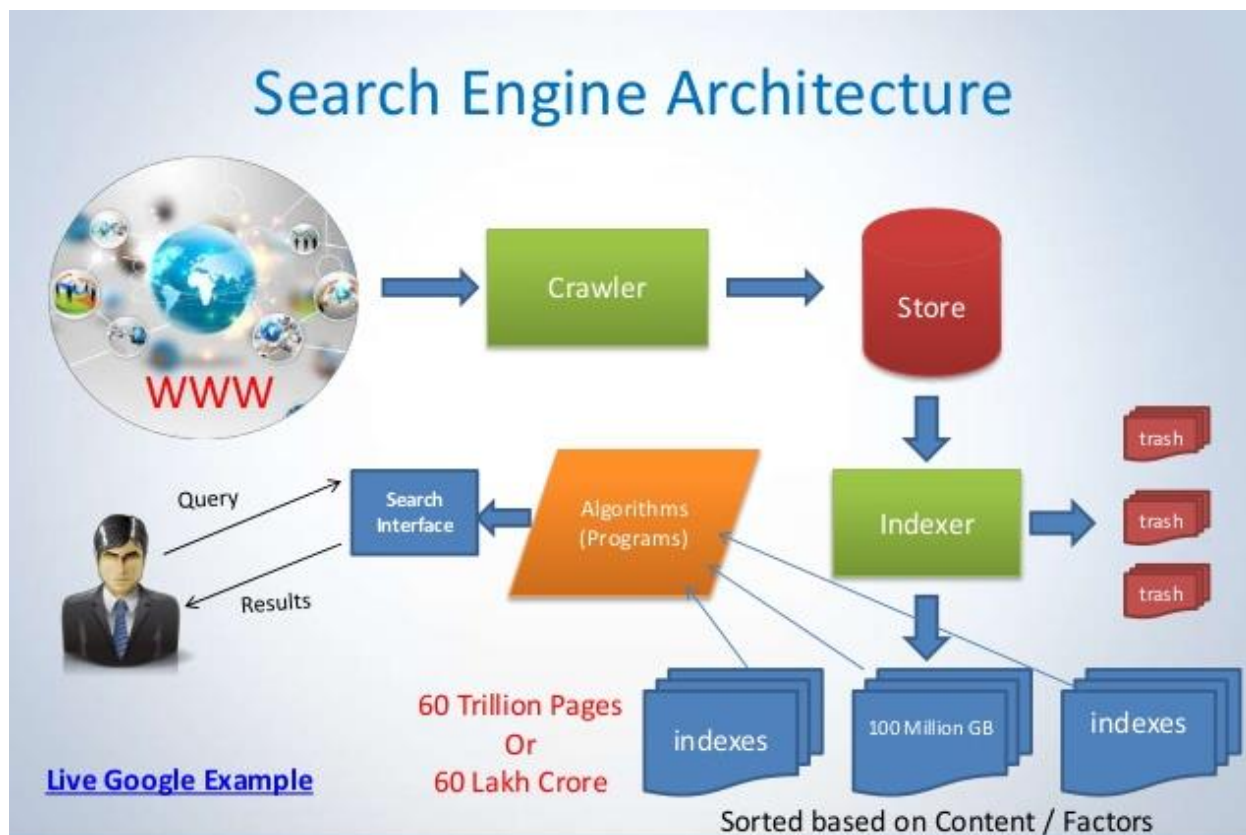
## انبارہ صفحات (Page Repository)

صفحاتی که توسط ماجول درون کاو از گستره جهانی وب استخراج می شوند در یک بانک اطلاعاتی به نام انبارہ صفحات ذخیره می شود تا به منظور شاخص دهی و تحلیل محتوایی در اختیار دیگر ماجول ها قرار بگیرد. ماجول درون کاو، عملیات استخراج و ذخیره صفحات به اتمام برسد یا آنکه تمام آدرس هایی که ماجول کنترل درون کاو در اختیار دارد ملاقات و استخراج شوند. تفاوت موتورهای جستجوی مختلف در الگوریتم کنترل درون کاو نمود بیشتری پیدا می کند. در برخی از این الگوریتم ها عملیات جستجو و استخراج صفحات وب به صورت غیر منقطع آنقدر ادامه می یابد تا هیچ صفحه ملاقات نشده ای باقی نماند. در این الگوریتم ها عملیات ملاقات و استخراج صفحات، از یک سایت وب با آدرس خاص، آغاز و به تمام آدرس های دیگری نیز که در این صفحات درج شده اند تسری (سرایت) می یابد. در برخی دیگر از الگوریتم ها عملیات ملاقات و استخراج صفحات یک وب سایت با آدرس خاص، به وب سایت هایی با پسوند نام حوزه (مانند com) متفاوت تسری نخواهد یافت. در گونه دیگری از الگوریتم های درون کاو، ملاقات و استخراج صفحات بین حوزه های سطح بالا محدود می ماند. به عبارت بهتر وقتی فرآیند استخراج صفحات حوزه com. در حال انجام است، ملاقات پسوند های net، org، و غیره در دستور کار قرار نمی گیرد.



## ماجول شاخص دهی (Indexer)

این ماجول یکایک صفحات موجود در در انباره صفحات را مرور کرده و کلمات موجود در آنها را پس از استخراج، به همراه URL مربوطه در یک جدول بسیار عظیم درج می کند. نتیجه این ماجول جدولی است که مشخص می کند هر کلمه در کدام صفحه و با چه آدرسی و در کجای صفحه ظاهر شده است.



یکی دیگر از خروجی های ماجول شاخص دهی، بانک اطلاعاتی "شاخص ساختاری" یا structure index است. این بانک اطلاعاتی چگونگی پیوند خوردن صفحات به یکدیگر را منعکس می کند. بعبارت دیگر گراف حاصل از صفحات و لینک هایی که در آنها به یکدیگر داده شده، در این شاخص ساختاری درج شده است. بدیهی است که این گراف منعکس کننده مسیری که ماجول درون کاو بر اساس آن صفحات وب را ملاقات و استخراج کرده نیز هست.

### Collection Analysis Module

از آنجاکه ماهیت ارتباط بین صفحات به غیر از ارجاعات و لینک هایی که مستقیماً باهم دارند به طرز پیچیده ای در حال تغییر است لذا این ماجول مسول تحلیل های محتوایی و هوشمند مجموعه صفحات ذخیره شده در انباره صفحات است.

## Utility Index

این شاخص می تواند خروجی ماجول قبلی باشد که پس از تحلیل کل مجموعه انباره صفحات بدست می آید. این شاخص ها می توانند بسیار متنوع و متعدد باشند. بعنوان مثال یکی از این شاخص ها می تواند تعداد تصاویر موجود در هر صفحه، تعداد لینک ها، رتبه اقتصادی وب سایت صاحب آن صفحه و دهها شاخص ریز و درشت دیگر باشد.

### موتور پرس و جو

این ماجول موظف است درخواست های جستجو را از کاربران دریافت و آنها را در قالبی مناسب جهت شروع فرآیند جستجو سازماندهی کند. بدیهی است که اغلب کاربران جستجوی موضوع مورد نظر خود را با یک یا دو کلمه ی کلیدی آغاز می کنند و طبعا نتیجه جستجو مجموعه بسیار بزرگی از آدرس ها خواهد بود. آغاز این مجموعه باید به نحو مناسبی رتبه دهی شود تا آیتم های مهم تر و بسیار مرتبط با درخواست کاربر در صدر نتایج جستجو قرار بگیرد. موتور جستجو به شدت به شاخص های ایجاد شده و همچنین انباره صفحات متکی است.

### ماجول رتبه بندی

این ماجول موظف است نتایج حاصل از جستجوی کلمات کلیدی کاربر را به نحوی رتبه دهی کند تا مرتبط ترین آیتم در صدر نتایج قرار بگیرد. هرچه این ماجول هوشمندانه تر و دقیق تر جواب ها را غربال و رتبه دهی کند، رضایت کاربران را بیشتر فراهم خواهد کرد.

### درون کاوی صفحات وب

ماجول درون کاوی موظف است صفحات وب را برای تحلیل و ایجاد شاخص، در روالی منظم، سریع و جامع استخراج کرده و تحویل انباره صفحات بدهد. این ماجول با یک مجموعه اولیه از url ها به نام SO آغاز به کار می کند. این URL ها در قالبی مناسب درون یک صف اولویت دار قرار می گیرند. ماجول درون کاو از سر این صف آدرسی را استخراج کرده و همانند یک کاربر معمولی تقاضای دریافت صفحه را ارسال و پس از آن، آدرس لینک ها و ارجاعات درون آن را بازیابی می کند تا پس از تحویل به ماجول کنترل درون کاو و اولویت

دهی در انتهای صف آدرس های URL درج شود. بدین ترتیب با حذف یک آدرس از سر این صف چندین آدرس جدید به انتهای آن افزوده می شود. ماجول کنترل درون کاو آدرس هایی که آخرین ملاقات شده اند را از میان آدرس ها حذف می کند تا در هر فرآیند درون کاوی هیچ صفحه ای بیش از یکبار ملاقات و بازیابی نشود. این فرآیند در ابتدا با رشد انفجاری طول صف آدرس ها مواجه خواهد شد نهایتا با تصمیم گیری و دخالت ماجول کنترل درون کاو خاتمه می یابد.

نظر به ماهیت دائما در حال تغییر وب و حجم عظیم صفحات، کنترل فرآیند درو کاوی وب با پیچیدگی های زیر روبروست:

## انتخاب صفحات

### معیار مبتنی بر گرایشات کاربران

فرض کنید مجموعه بزرگی از درخواست های جستجو که توسط کاربران یا گروهی خاص از کاربران ارسال شده را در اختیار داشته باشیم. این مجموعه از درخواست های جستجو را Q بنامید. هر یک از درخواست ها به نوبه خود از یک یا چند کلمه کلیدی تشکیل شده است. مجموعه کلید کلمات کلیدی که تمام کاربران در مجموعه درخواست هایشان جستجو کرده اند یک دیکشنری را تشکیل خواهد داد. بعنوان مثال دیکشنری شما دارای ده هزار کلمه است. از میان چهار میلیون کلمه ی کلیدی جستجو شده توسط یک گروه از کاربران کلمه IRAN مثلا ده هزار بار جستجو شده است. فراوانی نسبی این کلمه در بردار Q معادل  $1/400$  است. برای بقیه کلمات فراوانی نسبی به همین ترتیب محاسبه می شود.

### معیار مبتنی بر شهرت صفحات

شاید شهرت و اعتبار معیار بسیار خوبی برای تخمین میزان اهمیت یک صفحه وب باشد. حال سوالی که پیش می آید آن است که شهرت و اعتبار چگونه قابل اندازه گیری است؟ یکی از راه های اندازه گیری شهرت یک صفحه یا یک آیتم آن است که ببینیم در صفحات دیگر به چه تعداد به آن صفحه یا آیتم ارجاع شده است. به عبارت دقیق تر برای صفحه مفروض P، باید تعداد صفحاتی را که در کل گستره وب حداقل یکبار به P ارجاع داشته اند بشماریم. بدیهی است که هرچه شهرت و اعتبار یک صفحه یا سند بیشتر باشد تعداد وب سایت هایی که به اشکال مختلف به آن اشاره می کنند افزون تر است. به تعداد صفحاتی که به صفحه مفروض P ارجاع داشته اند اصطلاحا

**Backlink Count** (تعداد لینک های برگشتی) می گویند. بدیهی است که برای ارزیابی صحیح میزان اهمیت صفحات بایستی کل صفحات موجود در وب دریافت و لینک های درون آنها استخراج و شمارش شود.

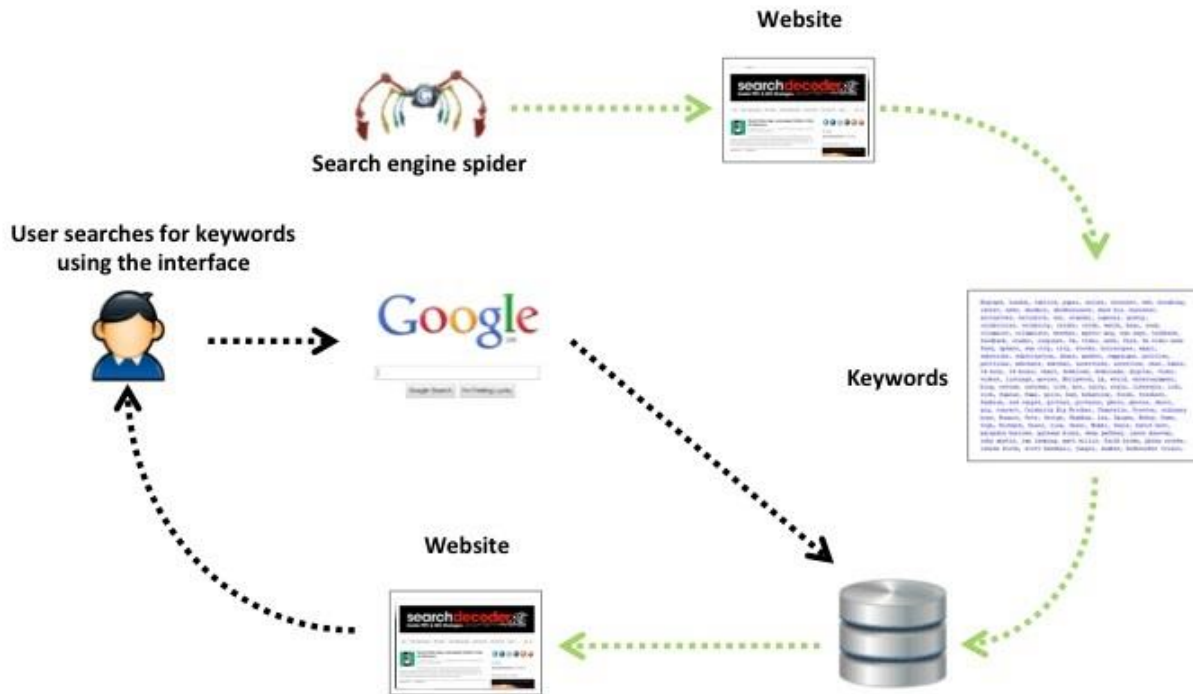
### معیار مبتنی بر محل قرار گرفتن صفحات

معیار قرار گرفتن یک صفحه در وب نیز می تواند معیار اهمیت آن باشد. منظور از محل قرار گرفتن صفحه، آدرس صفحه، ماهیت آدرس (از لحاظ `.net .com .edu` و ...) میزان فاصله آن از صفحه خانگی آن سایت است. این معیار تا حدودی می تواند اهمیت صفحه را تعیین کند بویژه وقتی که آماری از میزان مقبولیت وب سایت های جهان در اختیار باشد. معمولا این معیار را به تنهایی به کار نمی برند چه که میزان مقبولیت یک وبسایت که میزان اهمیت صفحات آن را تعیین می کند تابع تبلیغات و صرف هزینه است و تبلیغات، ارزش محتوایی ایجاد نمی کند.

### مدل درون کاوی

آنچه را که تاکنون بحث کردیم در خصوص معیار استخراج و ذخیره صفحات وب از پهنه اینترنت بود. "مدل درون کاوی" چالش دیگری است که موتور جستجو با آن سر و کار دارد و عبارتند از چگونگی آغاز و ختم فرآیند استخراج و ذخیره سازی صفحات وب از گستره بسیار عظیم وب می باشد. در این خصوص دو الگوی کلی قابل ارائه می باشد.





### الگوی کاوش و توقف (Crawl & Stop)

در این الگو، درون کاو، توسط ماجول کنترل کاو راه اندازی می شود و کارش را با آدرس یک صفحه مفروض  $P$  آغاز کرده و پس از ملاقات و دریافت دقیقاً  $K$  صفحه ی وب متوقف می شود.  $K$  عددی ثابت است و  $P_k$  صفحاتی هستند که به ترتیب از  $P_1$  تا  $P_k$  دارای بیشترین اهمیت بوده اند. یعنی  $P_1$  و به همین ترتیب که  $P_k$  که کمترین اهمیت را داراست.

### الگوی کاوش و توقف مبتنی بر مقدار آستانه (Crawl & Stop With Threshold)

در این الگو درون کاو  $C$  کارش را با آدرس صفحه مفروض  $P$  آغاز کرده و پس از دریافت  $K$  صفحه وب متوقف می شود ولیکن برخلاف مدل قبلی صفحاتی دریافت و ذخیره می شوند که پارامتر کمی اهمیت آنها از مقدار آستانه  $T$  بیشتر باشد.

## انبار ذخیره سازی

صفحاتی را که ماجول درون کاو ملاقات و استخراج می کند، درون یک سیستم ذخیره سازی بسیار عظیم درج می شوند تا برای شاخص دهی، رتبه بندی و عملیات پردازشی دیگر در اختیار باشند. حجم وحشتناک و ناهمگونی بسیار زیاد اسناد و صفحات دریافتی از وب، باعث شده که انبار ذخیره سازی صفحات وب، شباهت آنچنانی با پایگاه های اطلاعاتی رایج نداشته باشد. سیستم ذخیره سازی اطلاعات در انبار صفحات هر ساختاری که داشته باشد با چالش های زیر روبرو است:

### گسترش پذیری یا Scalability تا بی نهایت

بدیهی است که حجم صدها هزاران ترابایتی وب را نمی توان توسط یک کامپیوتر واحد سازماندهی و مدیریت کرد. لذا سیستم ذخیره سازی اطلاعات باید به صورت توزیع شده بر روی صدها تا هزاران کامپیوتر و توزیع شده باشد.

### پشتیبانی از دسترسی همزمان و دوگانه

انبار صفحات باید از دو روش دسترسی پشتیبانی کند: اولین روش، دسترسی مستقیم و تصادفی است تا بتوان به سرعت و مستقیماً یک صفحه وب خاص را بازیابی کرد. دومین روش، دسترسی جویباری است تا بتوان یک مجموعه بزرگ از صفحات را به یکباره بازیابی کرد. عموماً دسترسی مستقیم برای زمانی است که می خواهیم نسخه ای از سند ذخیره شده را تحویل کاربر نهایی بدهیم. دسترسی جویباری برای ماجول شاخص دهی یا ماجول های مشابه، نیاز است.

### بهنگام سازی عظیم و توده ای

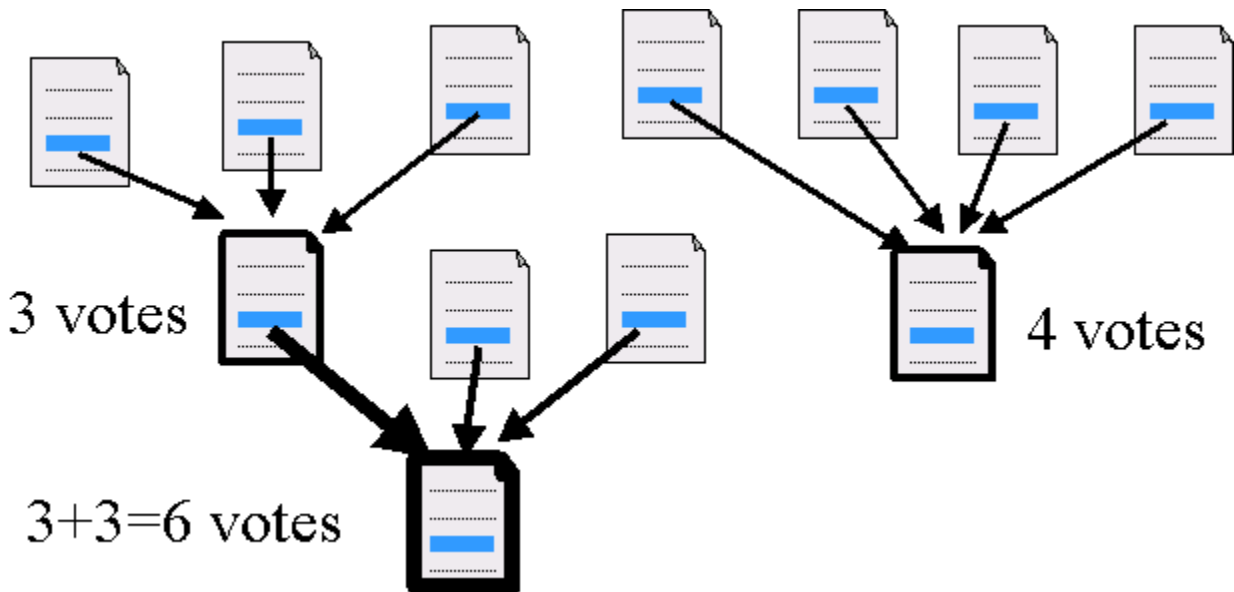
از آنجا که بخش بزرگی از صفحات وب پیوسته در حال تغییرند، لذا کل صفحاتی که در فرآیند تازه سازی دریافت می شوند بایستی جانشین صفحات قدیمی شوند. بنابراین سیستم ذخیره سازی باید اجازه بدهد که همزمان با دسترسی ماجول های دیگر به انبار صفحات، عملیات بهنگام سازی نیز در جریان باشد و هیچ تضادی بین این دو عمل همزمان وجود نداشته باشد.

### صفحات منسوخ

بدیهی است که بسیاری از صفحات وب برای همیشه از روی وب سایت ها برداشته می شوند لذا باید در فرآیند تازه سازی صفحات مکانیزمی اندیشیده شود تا این صفحات منسوخ از درون انبار ی ذخیره سازی صفحات نیز حذف گردد وگرنه حجم انبار با صفحات بی ارزش رو به رشد خواهد گذاشت.

## رتبه دهی و تحلیل لینک

موتور پرس و جو کلید واژه های کاربر را گرفته و با مراجعه به شاخص های موجود، صفحاتی را که با کلمات مورد نظر کاربر مطابقت دارند پیدا می کند. سوال اینجاست که آدرس صفحاتی که حاوی کلمات کلیدی مورد نظر کاربر هستند به چه ترتیبی تحویل وی داده می شوند. آیا تمام این صفحات از ارزش محتوایی یکسانی برخوردارند؟ آیا همه اینها مرتبط با موضوع مدنظر کاربر هستند؟



اینجاست که پای ماجول دیگری برای غربال سازی نتایج جستجو و رتبه بندی، در معماری موتور جستجو باز می شود. ماجول رتبه بندی یا Ranking Module در تمام موتورهای جستجو است که پس از غربال کردن نتایج بی ارزش یا کم ارزش، آنها را بر حسب اهمیت شان رتبه بندی مرتب می کند تا آنچه که کاربر دریافت می دارد فهرست مرتب شده ای از صفحات مرتبط با کلید واژه هایش باشد.

برای رتبه بندی محتوای درون صفحات موارد زیر را در نظر می گیرد:

- دفعات تکرار کلمات کلیدی
- ترتیب و مجاورت کلمات کلیدی
- محل درج کلمات کلیدی
- درج کلمات کلیدی درون آدرس صفحه

- پر رنگ بودن کلمات کلیدی و اندازه قلم
- بهره گرفتن از برچسب های توصیفی متا یا همان Meta Description tag
- بهره گیری از برچسب های Alt Tag

### فایل robot.txt و کاربرد برچسب Alt tag

این فایل یک فایل متنی است که بر روی سرویس دهنده وب و درون دایرکتوری اصلی هر وب سایت ذخیره می شود و تنظیمات گردش و سرکشی به اعماق آن وب سایت را عرضه کرده است. در حقیقت این فایل زحمت درون کاو را کاهش خواهد داد. این فایل با خطوط زیر آغاز می شود:

User-agent:

Disallow:

در بخش اول نام برنامه راهنمای ب سایت مشخص می شود. و در بخش دوم نام فایل ها یا دایرکتوری هایی که نباید توسط درون کاو ها بازدید و یا در نتایج جستجو ظاهر شوند، فهرست می شود. اگر کسی نخواهد هیچ نقطه ای از وب سایت او در فهرست جستجو قرار بگیرد و بازدید شود می تواند دو خط زیر را در این فایل درج کند:

User-agent:\*

Disallow: /

**\*\* تمرین: خصوصیت Alt tag در html بررسی شود.**

### روش های معنا بخشیدن به صفحات وب

#### XML

در ابتدا باید بدانیم XML برای شرح و بیان اطلاعات طراحی شده است که در نمایه و سند سازی هم کاربرد دارد و HTML برای نمایش دستورات نوشته شده استفاده می شود. لازم است بدانید که XML يك Markup Language مانند HTML است.

زبان های نشانه گذاری، مانند HTML یا XML ، ابزار مورد استفاده برای شناسایی ساختارهای مختلف در یک صفحه وب می باشند. هار دلیل اساسی برای استفاده از xml به عنوان ابزاری در طراحی وب سایت ارائه شده است:

۱. سهولت: یادگیری XML آسان است زیرا شما بر طرح و ساختار صفحه کنترل دارید. برنامه نویسان در طراحی سایت ، نام عنصر مورد نظر خود را توسعه می دهند. این به این معنی است که هیچ محدودیتی برای تعداد برچسب هایی است که می تواند بر روی یک سند XML استفاده شود.
۲. متن قابل خواندن XML: می تواند از طریق اکثر مرورگرها و داده استخراج شده موثر منقطع شود. از آنجا که طراحی سایت یا به عبارت صحیح تر، طراحان سایت تمام ویژگی ها، برچسب ها و ساختار عناصر را خلق می کنند، XML می تواند دارای داده های متا و برچسب های قابل تشخیص SEO در هر نقطه باشند.
۳. محتوا XML: یک ظرف ذخیره سازی است. که باعث می شود محتوای صفحه از ظاهر آن با ارزش تر شود XML. بر روی داده ها تمرکز دارد نه ویژگی های چگونگی نوشتن کدها
۴. کمک به طراحی سایت

## URL

مخفف Uniform Resource Locator می باشد و به معنی "مشخص کننده ی موقعیت منبع بصورت یک شکل" است که یک نوع URI به حساب می آید. منظور از URL همان آدرس عکس ها و ویدئو ها و صفحات روی اینترنت است.

## URI

مخفف Uniform Resource Identifier و به معنی "شناسه یک شکل منابع" و شناسه یا آدرسی است که از روی اجزای آن می توان اطلاعات کافی در مورد آن منبع از جمله رده یک شی و رده های بالاتر آن بدست آورد.

مثالی از یک URL بدون بکارگیری مفاهیم معنایی:

<http://aftab.cc/linkestan/showlink.php?id=1400>

از روی این آدرس نمیتوان فهمید که این لینک در مورد چه چیزی است. همان لینک با بکارگیری مفاهیم وب معنایی:

<http://aftab.cc/linkestan/link/pdf/learn-c-sharp-in-24-hours>

از روی این لینک میتوان فهمید که لینک به یک pdf در مورد آموزش C# است.

## RDF

مخفف Resource Description Framework و به معنی "چهارچوب توصیف منابع" است. زبانی

است که به لغات یک صفحه برای موتورهای جستجو معنا می بخشد و ارتباط بین لغات را مشخص می کند.

با استفاده از XML می توانیم موجودیت های مختلف و صفات آنها را تگ گذاری کنیم اما معنای آنها را نمی

توانیم بیان کنیم. بعنوان مثال اگر کسی برچسب قیمت را با PRICE نشانه گذاری کند و شخص دیگری با COST

نشانه گذاری کند هیچ راهی برای فهم مترادف بودن این دو عبارت به وسیله XML وجود ندارد، اینجاست که

RDF به کار می آید.

هر عبارت RDF بصورت یک سه تایی بیان می شود: فاعل، گزاره، مفعول

## نمایه

بعبارتی نمایه لیستی نظام یافته برای هدایت منظم یک متن، محتوا، مجموعه ای از مدارک یا هرگونه اطلاعات صبط شده است. نمایه، استفاده کنندگان را از اطلاعات معلوم به اطلاعات اضافی ناشناخته هدایت می کند.

## دلایل نیاز به نمایه سازی

۱. تحلیل موثر و سریع اسناد
۲. افزایش حجم اطلاعات
۳. سرعت پایین بازیابی اطلاعات در روش های کتابداری سنتی
۴. تغییر روش های ذخیره و بازیابی اطلاعات

## اهداف نمایه سازی

۱. زمان و تلاش برای یافتن اطلاعات را به حداقل میزان کاهش می دهد.
۲. موفقیت جستجوی استفاده کننده را به حداکثر می رساند.
۳. برقراری ارتباط میان مفاهیم
۴. تنظیم شناسه ها به ترتیبی نظام مند و موثر
۵. شناسایی سریع مدارک در یک مجموعه
۶. سازماندهی اطلاعات به قصد بازیابی سریع و آسان مدارک

## فرا داده یا Metadata

به اطلاعات اضافه که درباره یک داده ثبت می شود تا در طبقه بندی از آن ها استفاده شود گفته می شود. مانند اطلاعات متنوع یک عکس از جمله زمان عکاسی، اندازه، ابعاد، موقعیت جغرافیایی و ... که به نوعی اطلاعات اضافه گفته شده و به سیستم عامل جهت انجام اموراتی همچون دسته بندی استفاده می شود. همچنین در طراحی وب سایت به اطلاعاتی که در تگ meta در تگ head قرار داده می شود که به موتورهای جستجو در دسته بندی سایت ها کمک می کند.

## معرفی برخی متا تگ ها در طراحی وب

**<meta name="description" content="">**

این متا تگ حاوی توضیح کوتاهی درباره هدف سایت است که معمولاً در نتایج جستجو در زیر عنوان هر نتیجه

نمایش داده میشود.

**<meta name="keywords" content="keyword1,keyword2,...">**

حاوی نهایتاً ده کلمه کلیدی درباره هر صفحه وب. کلماتی که بیشترین ارتباط را با موضوع مطالب صفحه دارند و با کما جدا شده اند.

**<meta name="author" content="Hamid Reza Niroomand">**

حاوی نام نویسنده مقاله موجود در صفحه جاری

**<meta name="rating" content="general">**

حاوی رده سنی مرتبط با محتوای صفحه

**<meta name="generator" content="PHP-Nuke">**

نام cms یا سیستم مدیریت که توسط آن، سایت ایجاد شده است.

**<meta name="distribution" content="global">**

حاوی منطقه جغرافیایی مرتبط با محتوای صفحه

## زبان های نمایه سازی

زبان نمایه سازی زبانی است که برای توصیف موضوع یا سایر جنبه های اطلاعات یا مدارک در نمایه به کار می رود. زبان نمایه سازی به زبان ساختگی و قراردادی اطلاق می شود که برای مقاصد نمایه سازی به ویژه قابلیت بازیابی اطلاعات و مدارک به کار گرفته می شود.

شاید بتوان گفت که مهم ترین موضوع در نمایه سازی، انتخاب "کلید واژه" به عنوان مدخل است چرا که همه واژه های یک مقاله یا عنوان آن مقاله از نظر نویسنده و جوینده نقش کلیدی ندارند.

## برخی از زبان های نمایه سازی

۱. زبان آزاد



در این زبان نمایه ساز آزاد است از هر واژه ای که بتواند اطلاعات مدرک را برساند برای انتخاب به عنوان معیار قرار گیرد. بعنوان مثال اگر یک سند حاوی جمله "امروزه علم کامپیوتر یکی از کاربردی ترین علوم به حساب می آید" باشد. کلماتی مانند: علم (که در این سند موجود است)، دانش(که در سند موجود نیست)، کامپیوتر(که در سند موجود است)، رایانه(که در سند موجود نیست) و ...

۲. زبان طبیعی

اگر کلماتی که نمایه ساز انتخاب کند در سند موجود باشد این نمایه سازی به زبان طبیعی مشهور است. مثال اگر یک سند حاوی "امروزه علم کامپیوتر یکی از کاربردی ترین علوم به حساب می آید" باشد، فقط کلماتی مانند علم، کامپیوتر، کاربردی می توانند در نظر گرفته شوند.

۳. زبان مقید یا مهار شده

در این زبان نمایه ساز بعد از درک مدارک و موضوع باید از بین واژه ها و اصطلاحاتی که از قبل تهیه و اعتبار سنجی و وضع شده مناسب ترین واژه را برای موضوع مدرک انتخاب کند. بنابراین نمایه ساز فقط قدرت انتخاب دارد و نه بیشتر و دیگر آن آزادی عمل را در گزینش واژه ندارد. مثال اگر یک سند حاوی جمله "امروزه علم کامپیوتر یکی از کاربردی ترین علوم به حساب می آید" باشد و نمایه ساز مقید شده باشد که فقط از کلمات کاملاً فارسی کمک بگیرد، فقط کلماتی مانند: دانش، رایانه، کاربردی می توانند مورد استفاده قرار بگیرند.

## انواع نمایه ها از لحاظ روش تنظیم

۱. نمایه های الفبایی

مدخل ها به ترتیب حروف الفبا چیده می شوند و اکثر نمایه های کنونی به این صورت است.

۲. نمایه های زمانی

مدخل ها به ترتیب زمان از قدیم به حال مرتب شده و بیشتر برای کتب تاریخی به کار می رود.

۳. رده ای یا موضوعی

مدخل ها بر اساس عنوان موضوعات مرتب می شوند. بیشتر

۴. نمایه تکاملی

مدخل ها به ترتیب تکامل چیده شده اند و بیشتر برای متون و کتب زمین شناسی مناسب است.

این نوع نمایه شامل یه فهرست فرعی تحت هر یک از مقالات منتشر شده است که حاوی مقالاتی است که به آن مقالات استناد کرده اند. به عبارت دیگر در مورد یک مقاله خاص، نمایه استنادی مشخص می کند که آن مقاله توسط چه مقالات دیگری که بعد از آن نوشته شده مورد استفاده قرار گرفته اند.

### انواع نمایه سازی از لحاظ روش تهیه

- نمایه سازی دستی

در این نوع نمایه سازی، مدخل ها و حتی ارجاعات توسط انسان و به صورت دستی انتخاب می شوند.

- نمایه سازی خودکار

فرآیند استخراج مجموعه ای از مدخل ها ی نمایه ای بیانگر موضوع متن هستند توسط رایانه را نمایه سازی خودکار می گویند.

### دلایل نیاز به نمایه سازی خودکار

- افزایش حجم اطلاعات موجود
- تنوع ساختار در قالب هایی مانند مقاله، پایان نامه رساله و...
- نیاز به خلاصه سازی مقالات چکیده و ضرورت یکسان سازی قالب آن ها
- ارائه درخواست های اطلاعاتی بهره گیران به صورت اصطلاح یا واژه های منفرد.

### آشنایی با برخی نمایه های قدیمی

#### نمایه های گردان

۱. نمایه کوئیک KWIC

مخفف **Key Word In Context** و به معنی "کلمه کلیدی، در متن" است. نوعی نمایه گردان است که هر بار یکی از کلمات عنوان به صورت چرخشی به عنوان مدخل قرار می گیرد. همچنین به آن در اصطلاح نمایه درون بافتی نیز می گویند چون مدخل جزئی از بافت کلی است. این نمایه یکی از موفق ترین نمایه های ماشینی است.

مثال: فرض کنید ۴ مقاله با عنوان زیر داشته باشیم که شماره بازیابی آن ها به ترتیب ۱۰۱، ۱۰۲، ۱۰۳، ۱۰۴ است با روش کوئیک یک نمایه نامه برای آن ها تهیه کنید.

۱۰۱ اقتصاد مسکن در ایران

۱۰۲ بررسی آثار اقتصادی رکود بازار فرش

۱۰۳ نقش فرش در اقتصاد ایران

۱۰۴ عوامل اقتصادی مهاجرت از روستاها به شهرها

ارجاع	بافت		
	بخش اول	مدخل	بخش دوم
۱۰۳	نقش فرش در	اقتصاد	ایران
۱۰۱		اقتصاد	مسکن در ایران
۱۰۲	بررسی آثار	اقتصادی	رکود بازار فرش
۱۰۴	عوامل	اقتصادی	مهاجرت از روستاها به شهرها
۱۰۱	اقتصاد مسکن در	ایران	
۱۰۳	نقش فرش در اقتصاد	ایران	
۱۰۲	بررسی آثار اقتصادی رکود	بازار	فرش
۱۰۲	بررسی آثار اقتصادی	رکود	بازار فرش
۱۰۴	عوامل اقتصادی مهاجرت از	روستاها	به شهرها
۱۰۴	عوامل اقتصادی مهاجرت از روستاها به	شهرها	
۱۰۲	بررسی آثار اقتصادی رکود بازار	فرش	
۱۰۳	نقش	فرش	در اقتصاد ایران
۱۰۱	اقتصاد	مسکن	در ایران
۱۰۴	عوامل اقتصادی	مهاجرت	از روستاها به شهرها
۱۰۳		نقش	فرش در اقتصاد ایران

۲. نمایه کووک KWOC

مخفف **Key Word Out of Context** و به معنی "کلمه کلیدی، خارج از متن" است. نوعی نمایه گردان است هر بار یکی از کلمات عنوان به ترتیب از عنوان خود خارج شده و مقدم بر سایر اجزای آن قرار می گیرد. سپس عنوان به ترتیب طبیعی خود در زیر این واژه یا به دنبال آن می آید. به آن در اصطلاح نمایه برون بافتی نیز می گویند چون مدخل از درون بافت خود خارج شده و مقدم بر سایر اجزا قرار می گیرد. این نمایه برای جبران مسائل ناشی از محدود بودن جا، کوتاه شدن عنوان و نیز مشکلات خواندن نمایه ی درون بافتی طراحی شده است.

مثال زیر جزئیات دقیق تری ارائه می دهد:

ارجاع	بافت	مدخل
101	اقتصاد مسکن در ایران	اقتصاد
103	نقش فرش در اقتصاد ایران	اقتصاد
102	بررسی آثار اقتصادی رکود بازار فرش	اقتصادی
104	عوامل اقتصادی مهاجرت از روستاها به شهرها	اقتصادی
101	اقتصاد مسکن در ایران	ایران
103	نقش فرش در اقتصاد ایران	ایران
102	بررسی آثار اقتصادی رکود بازار فرش	بازار
102	بررسی آثار اقتصادی رکود بازار فرش	رکود
104	عوامل اقتصادی مهاجرت از روستاها به شهرها	روستاها
104	عوامل اقتصادی مهاجرت از روستاها به شهرها	شهرها
102	بررسی آثار اقتصادی رکود بازار فرش	فرش
102	بررسی آثار اقتصادی رکود بازار فرش	فرش
101	اقتصاد مسکن در ایران	مسکن
104	عوامل اقتصادی مهاجرت از روستاها به شهرها	مهاجرت
103	نقش فرش در اقتصاد ایران	نقش

### ۳. نمایه جایگشتی Permuterm

اساس کار آن مانند کوئیک و کووک گردش کلید واژه های عنوان است. کلید واژه های عنوان به صورت دوتایی انتخاب می شوند و یک اصطلاح را تشکیل می دهند. گردش کار با دو کلید واژه است.

### قنایط ضعف و قوت نمایه های گردان

۱. در مقایسه با نمایه موضوعی بسیار ارزان و ساده و سریع تهیه می شود.
۲. نمایه گردان منعکس کننده واژگان رایج در زمینه مربوطه هستند.
۳. نمایه گردان با فراهم آوردن بافت نسبی تصویر روشن تری از موضوع می دهد.
۴. در تهیه نمایه، محتوای موضوعی مدرک توسط نمایه سازان مختلف تفسیر نمی شوند و به این جهت جایی برای اشتباه و اختلاف از لحاظ تعیین و تبیین باقی نمی ماند.
۵. گاهی طول مدخل هایی که با یک کلید واژه درست می شود به یک یا چند صفحه بالغ می گردد.

## اصطلاح مرجح چیست؟

اصطلاحی است که نسبت به مترادف هایش برتری (رجحان) دارد. مانند دانش به جای علم و یا رایانه به جای کامپیوتر

### ملاک انتخاب اصطلاح مرجح:

۱. رواج داشتن: مثل جامعه شناسی به جای علم الاجتماع
۲. بومی بودن: مثل آرمان گرایی به جای ایده آلیسم / رایانه به جای کامپیوتر
۳. جدید و امروزی بودن: مثل هواپیما به جای طیاره
۴. اختصار مشهور به جای عبارت کامل: یونسکو
۵. اجتناب از کلمات ترکیبی به جز در مواردی که اصطلاح قابل جداسازی نیست: مثل آبله مرغان

## ویژگی های یک نمایه خوب چیست؟

### ۱. صحت

هرچه خطاهای ارجاع به صفحه ها کاهش یابد و از مدخل های صحیح استفاده شود، نمایه کارایی بیشتری خواهد داشت.

### ۲. عمق نمایه سازی

اینکه در هر صفحه از سند چند کلمه به عنوان مدخل قرار گرفته، عمق نمایه سازی گفته می شود. عمق نمایه سازی باید با نیاز کاربران متناسب باشد. به عنوان نمونه، یک نمای متوسط ۳ تا ۵ مدخل اصلی برای هر صفحه از متن و یک نمای ضعیف ۱ تا ۲ مدخل دارد.

### ۳. ایجاز

هرچه از سر عنوان های فرعی به صورت فشرده و به هم بسته استفاده شود نمایه بهتری خواهیم داشت.

### ۴. ارجاعات متقابل

ارجاعات باید منطقی و مفید باشند.

### ۵. سر عنوان های منطقی

سرعنوان ها باید از میان گزینه های مختلف بهترین آن ها انتخاب شده باشند.

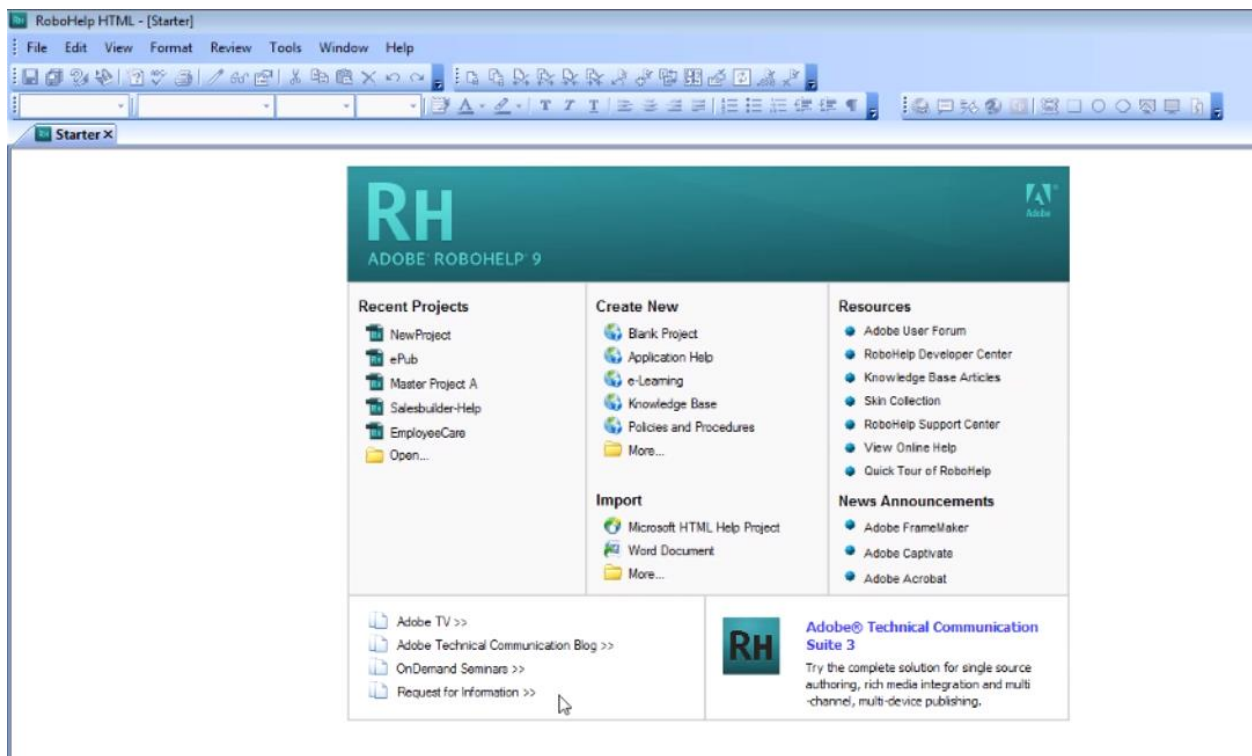
## نرم افزار های نمایه سازی

- نرم افزار HTML/Prep
- نرم افزار HTML Indexer
- نرم افزار RoboHelp HTML Edition

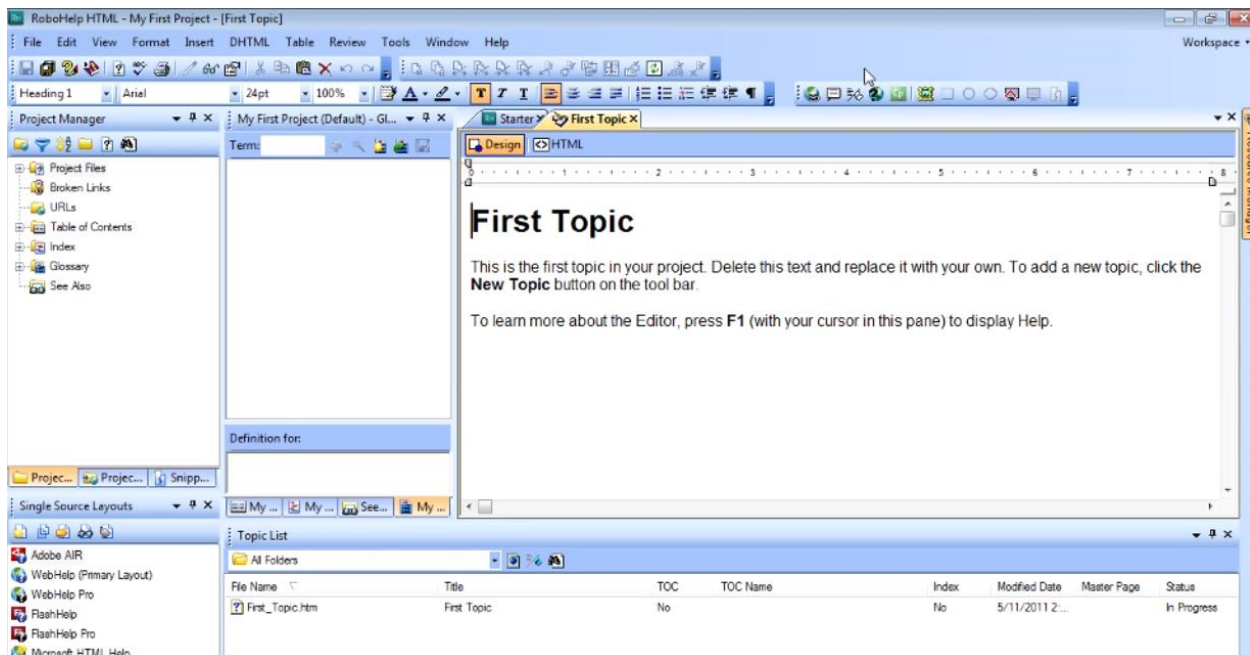
## آموزش نرم افزار RoboHelp HTML Edition

تصور کنید در حال کار با نرم افزار مورد علاقه خود هستید با فشار کلید F1 و باز شدن پنجره Help خواهید دید که نرم افزار RoboHelp این چنین محصولاتی را تولید می کند. در واقع این نرم افزار دارای امکانات لازم برای ساختن سیستم های Help استاندارد می باشد. با این نرم افزار م می توان فایل های pdf، WebHelp، FlashHelp، Adobe Air، HTML را تولید کرد.

شکل زیر پنجره شروع اولیه را نمایش می دهد که کاربر می تواند برای ورود به برنامه قسمت مورد نظر انتخاب کند.



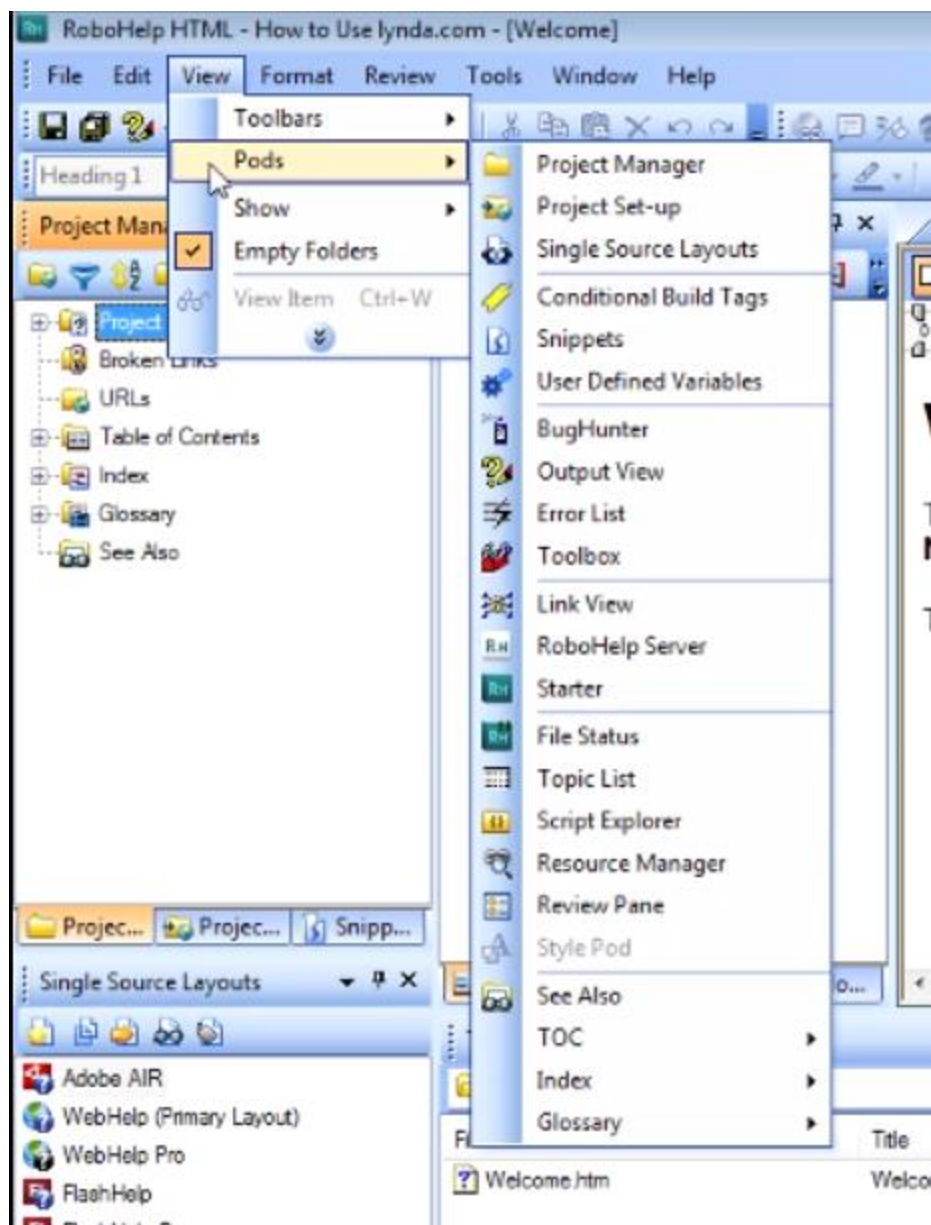
پس از انتخاب یک پروژه خام و ورود به برنامه صفحه کاری زیر را مشاهده خواهید کرد:



از طریق گزینه زیر می توانید نمونه قالب پروژه های از قبل آماده شده را باز کنید.

## File -> New -> Projects

در این نرم افزار ابزار های متنوعی در اختیار شما گذاشته می شوند که به آن ها Pods گفته می شوند و در پنل سمت چپ قرار می گیرند. همچنین می توانید به تعداد زیادی از آن ها در منوی زیر دست پیدا کنید.



پسوند فایل های این نرم افزار با xpj ذخیره می شوند.



## لیست پروژه های انتخابی

ردیف	عنوان	تعداد گروه های مجاز	حد اکثر تعداد هر گروه	توضیحات
1	آموزش کامل XML	سه	سه	حد اقل ۳۰ صفحه
2	آموزش کامل Microsoft Word	دو	سه	حد اقل ۳۰ صفحه
3	آموزش کامل نرم افزار Microsoft Excell	دو	سه	حد اقل ۳۰ صفحه
4	تهیه یک تحقیق کامل در مورد کتابخانه ملی	دو	سه	شامل تاریخچه، عضویت، نحوه مدیریت، منابع موجود، شیوه امانت دهی و سایر مطالب دیگر حد اقل ۲۰ صفحه

\*\*لینک دانلود نرم افزار RoboHelp HTML Edition

[/http://p30download.com/fa/entry/61807](http://p30download.com/fa/entry/61807)