

به نام خدا

سازنده: سهراب نیازی

وب سایت: [WwW.NiaziSoft.blogfa.CoM](http://WwW.NiaziSoft.blogfa.CoM)

ایمیل: [NiaziSoft\\_Help@Yahoo.CoM](mailto:NiaziSoft_Help@Yahoo.CoM)

موضوع: موتور جستجوگر چگونه کار میکند

موتور جستجوگر

الله اعلم

موتور جستجوگر چگونه کار می کند؟

وقتی جستجویی در یک موتور جستجوگر انجام و نتایج جستجو ارایه می شود، کاربران در واقع نتیجه کار بخش های متفاوت موتور جستجوگر را می بینند. موتور جستجوگر قبلاً " پایگاه داده اش را آماده کرده است و این گونه نیست که درست در همان لحظه جستجو، تمام وب را بگردد. بسیاری از خود می پرسند که چگونه امکان دارد گوگل در کمتر از یک ثانیه تمام سایت های وب را بگردد و میلیون ها صفحه را در نتایج جستجوی خود ارایه کند؟

نه گوگل و نه هیچ موتور جستجوگر دیگری توانایی انجام این کار را ندارند. همه آنها در زمان پاسخ گویی به کاربران، تنها در پایگاه داده ای که در اختیار دارند به جستجو می پردازند و نه در وب!

موتور جستجوگر به کمک بخش های متفاوت خود، اطلاعات مورد نیاز را قبلاً جمع آوری، تجزیه و تحلیل می کند و آنرا در پایگاه داده اش ذخیره می نماید و به هنگام جستجوی کاربر تنها در همین پایگاه داده می گردد.

بخش های مجزای یک موتور جستجوگر عبارتند از:

- Spider یا عنکبوت
- Crawler یا خزنده
- Indexer یا بایگانی کننده
- Database یا پایگاه داده
- Ranker یا سیستم رتبه بندی

## Spider - (عنكبوت)

اسپایدر یا روبات (Robot)، نرم افزاری است که کار جمع آوری اطلاعات مورد نیاز یک موتور جستجوگر را بر عهده دارد. اسپایدر به صفحات مختلف سر می زند، محتوای آنها را می خواند، اطلاعات مورد نیاز را جمع آوری می کند و آنها در اختیار سایر بخش های موتور جستجوگر قرار می دهد.

کار یک اسپایدر، بسیار شبیه کار کاربران وب است. همانطور که کاربران، صفحات مختلف را بازدید می کنند، اسپایدر هم درست این کار را انجام می دهد با این تفاوت که اسپایدر کدهای HTML صفحات را می بیند اما کاربران نتیجه حاصل از کنار هم قرار گرفتن این کدها را.

Index.html صفحه ای است که کاربران آنها به صورت شکل (1) می بینند:

اما یک اسپایدر آنها چگونه می بیند؟

برای این که شما هم بتوانید دنیای وب را از دیدگاه یک اسپایدر ببینید، کافی است که کدهای HTML صفحات را مشاهده کنید. برای این کار در مرورگر مورد استفاده خود، مسیر نشان داده شده در شکل (2) دنبال

کنید. (View Menu <--- Souce)

با انجام این کار فایل متنی شکل (3) به شما نشان داده می شود:

آیا این دنیای متنی برای شما جذاب است؟

اسپایدر، به هنگام مشاهده صفحات، از خود بر روی سرورها رد پا برجای می‌گذارد. شما اگر اجازه دسترسی به آمار دید و بازدیدهای صورت گرفته از یک سایت و اتفاقات انجام شده در آنرا داشته باشید، می‌توانید مشخص کنید که اسپایدر کدام یک از موتورهای جستجوگر صفحات سایت را مورد بازدید قرار داده‌اند.

یکی از فعالیتهای اصلی که در SEM انجام می‌شود تحلیل آمار همین دید و بازدیدها می‌باشد. اسپایدرها کاربردهای دیگری نیز دارند، به عنوان مثال عده‌ای از آنها به سایت‌های مختلف مراجعه می‌کنند و فقط به بررسی فعال بودن لینک‌های آنها می‌پردازند و یا به دنبال آدرس پست الکترونیکی (Email) می‌گردند.

#### - Crawler (خزنده)

کراولر، نرم‌افزاری است که به عنوان یک فرمانده برای اسپایدر عمل می‌کند. آن مشخص می‌کند که اسپایدر کدام صفحات را مورد بازدید قرار دهد. در واقع کراولر تصمیم می‌گیرد که کدام یک از لینک‌های صفحه‌ای که اسپایدر در حال حاضر در آن قرار دارد، دنبال شود. ممکن است همه آنها را دنبال کند، بعضی‌ها را دنبال کند و یا هیچ کدام را دنبال نکند.

کراولر، ممکن است قبلاً "برنامه ریزی شده باشد که آدرس‌های خاصی را طبق برنامه، در اختیار اسپایدر قرار دهد تا از آنها دیدن کند. دنبال کردن لینک‌های یک صفحه به این بستگی دارد که موتور جستجوگر چه حجمی از

اطلاعات یک سایت را می تواند در پایگاه داده اش ذخیره کند و همچنین ممکن است اجازه دسترسی به بعضی از صفحات به موتورهای جستجوگر داده نشده باشد.

شما به عنوان دارنده سایت، همان طور که دوست دارید موتورهای جستجوگر اطلاعات سایت شما را با خود ببرند، می توانید آنها را از بعضی از صفحات سایت تان دور کنید و اجازه دسترسی به محتوای آن صفحات را به آنها ندهید. تنظیم میزان دسترسی موتورهای جستجوگر به محتوای یک سایت توسط پروتکل Robots انجام می شود که در مقالات دیگر سایت به آن پرداخته شده است. به عمل کراولر ، خزش (Crawling) می گویند.

#### - Indexer (بایگانی کننده)

تمام اطلاعات جمع آورش شده توسط اسپایدر در اختیار ایندکسر قرار می گیرد. در این بخش اطلاعات ارسالی مورد تجزیه و تحلیل قرار می گیرند و به بخش های متفاوتی تقسیم می شوند. تجزیه و تحلیل بدین معنی است که مشخص می شود اطلاعات از کدام صفحه ارسال شده است، چه حجمی دارد، کلمات موجود در آن کدام است، کلمات چندبار تکرار شده است، کلمات در کجای صفحه قرار دارند و ...

در حقیقت ایندکسر، صفحه را به پارامترهای آن خرد می کند و تمام این پارامترها را به یک مقیاس عددی تبدیل می کند تا سیستم رتبه بندی بتواند پارامترهای صفحات مختلف را با هم مقایسه کند.

در زمان تجزیه و تحلیل اطلاعات، ایندکسر برای کاهش حجم داده ها از بعضی کلمات که بسیار رایج هستند صرفنظر می کند. کلماتی نظیر a, an, the, www, is و ... از این گونه کلمات هستند.

#### - Database (پایگاه داده)

تمام داده های تجزیه و تحلیل شده در ایندکسر، به پایگاه داده ارسال می گردد. در این بخش داده ها گروه بندی، کدگذاری و ذخیره می شود. همچنین داده ها قبل از آنکه ذخیره شوند، طبق تکنیکهای خاصی فشرده می شوند تا حجم کمی از پایگاه داده را اشغال کنند.

یک موتور جستجوگر باید پایگاه داده عظیمی داشته باشد و به طور مداوم حجم محتوای آنرا گسترش دهد و البته اطلاعات قدیمی را هم به روز رسانی نماید. بزرگی و به روز بودن پایگاه داده یک موتور جستجوگر برای آن امتیاز محسوب می گردد. یکی از تفاوتهای اصلی موتورهای جستجوگر در حجم پایگاه داده آنها و همچنین روش ذخیره سازی داده ها در پایگاه داده است.

در شکل (4) حجم پایگاه داده چند موتور جستجوگر با هم مقایسه شده است.

شکل 4- مقایسه حجم پایگاه داده چند موتور جستجوگر در دسامبر 2001

AllTheWeb =GG= Google, AV= Altavista, Fast

INK= Inktomi, NL= NorthernLight

## – Ranker (سیستم رتبه بندی)

بعد از آنکه تمام مراحل قبل انجام شد، موتور جستجوگر آماده پاسخ گویی به سوالات کاربران است. کاربران چند کلمه را در جعبه جستجوی (Search Box) آن وارد می کنند و سپس با فشردن **Enter** منتظر پاسخ می مانند.

برای پاسخگویی به درخواست کاربر، ابتدا تمام صفحات موجود در پایگاه داده که به موضوع جستجو شده، مرتبط هستند، مشخص می شوند. پس از آن سیستم رتبه بندی وارد عمل شده، آنها را از بیشترین ارتباط تا کمترین ارتباط مرتب می کند و به عنوان نتایج جستجو به کاربر نمایش می دهد.

حتی اگر موتور جستجوگر بهترین و کامل ترین پایگاه داده را داشته باشد اما نتواند پاسخ های مرتبطی را ارائه کند، یک موتور جستجوگر ضعیف خواهد بود. در حقیقت سیستم رتبه بندی قلب تپنده یک موتور جستجوگر است و تفاوت اصلی موتورهای جستجوگر در این بخش قرار دارد.

سیستم رتبه بندی برای پاسخ گویی به سوالات کاربران، پارامترهای بسیاری را در نظر می گیرد تا بتواند بهترین پاسخ ها را در اختیار آنها قرار دهد. حرفه ای های دنیای **SEM** به طور خلاصه از آن به **Algo** ( الگوریتم) یاد می کنند.

الگوریتم، مجموعه ای از دستورالعمل ها است که موتور جستجوگر با اعمال آنها بر پارامترهای صفحات موجود در پایگاه داده اش، تصمیم می گیرد که صفحات مرتبط را چگونه در نتایج جستجو مرتب کند. در حال حاضر قدرتمندترین سیستم رتبه بندی را گوگل در اختیار دارد.



می توان با ادغام کردن اسپایدر با کراولر و همچنین ایندکسر با پایگاه داده، موتور جستجوگر را شامل سه بخش زیر دانست که این گونه تقسیم بندی هم درست می باشد:

• کراولر

• بایگانی

• سیستم رتبه بندی

تذکر- برای سهولت در بیان مطالب بعدی هر گاه صحبت از بایگانی کردن (شدن) به میان می آید، مقصود این است که صفحه تجزیه و تحلیل شده و به پایگاه داده موتور جستجوگر وارد می شود.

برای آنکه تصور درستی از نحوه کار یک موتور جستجوگر داشته باشید داستان نامتعارف زیر را با هم بررسی می کنیم.

داستان ما یک شکارچی دارد. او تصمیم به شکار می گیرد:

- کار کراولر:

او قصد دارد برای شکار به منطقه حفاظت شده ابیورد، واقع در شهرستان درگز (شمالی ترین شهر خراسان بزرگ) برود.

## - پروتکل Robots :

ابتدا تمام محدودیت های موجود برای شکار در این منطقه را بررسی می کند:

- آیا در این منطقه می توان به شکار پرداخت؟

- کدام حیوانات را می توان شکار کرد؟

- حداکثر تعداد شکار چه میزانی است؟

- و ...

فرض می کنیم او مجوز شکار یک اورپال (نوعی آهو) را از شکاربانی منطقه دریافت می کند.

- کار اسپایدر

او اورپالی رعنا را شکار می کند و سپس آنرا با خود به منزل می برد.

- کار ایندکسر

شکار را تکه تکه کرده، گوشت، استخوان، دل و قلوه، کله پاچه و ... آنرا بسته بندی می کند و بخش های زاید

شکار را دور می ریزد.

- کار پایگاه داده

بسته های حاصل را درون فریزر قرار داده، ذخیره می کند.

- کار سیستم رتبه بندی

مهمانان سراغ او می آیند و همسر او بسته به ذائقه مهمانان برای آنها غذا طبخ می کند. ممکن است عده ای کله پاچه، عده ای آبگوشت، عده ای جگر و ... دوست داشته باشند. پخت غذا طبق سلیقه مهمانان کار سختی است. ممکن است همه آنها آبگوشت بخواهند اما آنها مسلماً "بامزه ترین آبگوشت را می خواهند!

نکته ها:

- شکارچی می توانست برای شکار کبک یا اورپال و یا هر دو به آن منطقه برود همانطور که موتور جستجوگر می تواند از سرور سایت شما انواع فایل (عکس، فایل متنی، فایل اجرایی و ...) درخواست کند.
- غذای خوشمزه را می توانید با نتایج جستجوی دقیق و مرتبط مقایسه کنید. اگر شکارچی بهترین شکار را با خود به منزل ببرد اما غذایی خوشمزه و مطابق سلیقه مهمانان طبخ نکرده، تمام زحمات هدر رفته است.
- به عنوان آخرین نکته این مقاله یاد آوری می کنم که به شکار اورپالی رعنا آن هم در منطقه حفاظت شده ابیورد، اصلاً فکر نکنید. اما توصیه می شود که حتماً از طبیعت بکر آن دیدن فرمایید (بدون اسلحه!).

## جستجو گر داخلی سایت



در صورتیکه پورتال و یا وب سایت شما دارای آنچنان وسعتی که نیاز به نصب و راه اندازی موتور جستجو در کارگزار اختصاصی، وجود ندارد، شما می توانید مشترک خدمات جستجوی سفارشی ما شوید .

موتور جستجوگر رادکام در رایانه های قدرتمند با توان پردازشی و ذخیره سازی بالا نصب و راه اندازی شده است. این رایانه ها با خطوط پر سرعت به شبکه جهانی وب متصل هستند. همه این کارها برای آن انجام شده است تا در کمترین زمان (کمتر از یک ثانیه) جستجوی واژگان مورد نظر شما انجام و به شما ارسال گردد.

رابط های برنامه نویسی (API) با زبانهای PHP, CSharp, Visual Basic, JSP طراحی شده است و در اختیاران قرار خواهد گرفت. شما متناسب با زبان برنامه نویسی مورد استفاده در وب سایت و یا پورتال خود، رابط مناسب را انتخاب و آن را نصب می کنید. این رابط های برنامه نویسی طوری پیاده شده اند که به راحتی با برنامه های شما یکپارچه شوند . کاربران وب سایت شما هیچگاه متوجه نخواهند شد که عملیات جستجو از خارج از محدوده وب سایت انجام می شود. چرا که هنگامی که کاربر جستجویی را آغاز می کند، عبارت مورد جستجو از طریق رابط برنامه نویسی جاسازی شده در برنامه های وب سایت، به کارگزار جستجو در شبکه رادکام ارسال می گردد. این کارگزار عبارت را جستجو کرده و نتیجه آن را در برای رابط برنامه نویسی در وب سایت شما ارسال می کند سپس این رابط نتیجه را برای تغییر و تحول و عرضه با شکل و قالب متناسب با وب سایت ، در اختیار شما قرار می دهد. همه این کارها در کمتر از یک ثانیه انجام می گردد. نمونه بارز این خدمت را در وب سایت رادکام

[www.radcom.ir](http://www.radcom.ir) می توانید مشاهده کنید. در این وب سایت رابط برنامه نویسی با زبان CSharp و در محیط dot Net مورد استفاده قرار گرفته است.

**Alltheweb:** یک موتور جستجوی خیلی بزرگ که اعتبار و عمومیت زیادی دارد. در سال 2002

اهراً یکی از مهمترین رقبای Google بشمار می رفت . در یک بررسی صورت گرفته توسط Pandecta magazine. که در چهارمین سه ماه سال 2002 انجام شده است ، تخمین زده شده است که alltheweb دومین و بزرگترین پایگاه داده را دارد ، همچنین در تست ارتباط نیز خوب کار میکند و بعد از گوگل و wisenut در رتبه سوم است. اما در تست سرعت معمولاً شکست میخورد.

**Altavista:** موتور جستجوی پیشگامی است که توسط overture در فوریه 2003 خریداری شده است.

**Ask jeeves:** یک موتور جستجوی نسبتاً عمومی است. ادعایش برای شهرت این است که به شما اجازه میدهد تا بجای استفاده تنها از کلمات کلیدی بتوانید درخواست خود را بصورت یک سئوالات متنی توضیحی وارد کنید Ask jeeves از teoma. overture و ODP نتایج جستجو را دریافت میکند .

**Excite:** یکی از موتورهای اصلی جستجو است.

**Google:** بزرگترین ، سریعترین و دقیقترین موتور جستجو است . گوگل با سیستم **pagerank** خودش مشهور است.

**Hotbot:** یک موتور جستجوی نسبتاً عمومی است ، اگرچه محبوبیت آن شدیداً با تسلط پیدا کردن گوگل کاهش یافته است **Hotbot** . یکبار گزارش شده است که بزرگترین پایگاه داده را در اختیار دارد. دریک بررسی صورت گرفته توسط **pandecta magazine** تخمین زده شده است که چهارمین پایگاه داده بزرگ را بعد از گوگل ، **alltheweb** و **wisenut** دارد .

**hotbot** به منظور رسیدن به سرعت از فن آوری محاسبه موازی بهره برداری میکند ( شبکه ایستگاه های کاری زمانی که شما قدرت محاسبه ترکیبی این مولفه های کوچکتر را اضافه میکنید، شما کارایی طبقاتی ابرکامپیوتر را به دست می آورید.

**Lycos:** خود را به عنوان یک موتور جستجو مطرح کرده است و در اواخر دهه 90 دارای رتبه بالایی از لحاظ محبوبیت بین کاربران اینترنتی بوده است.

**Overture:** بزرگترین و عمومی ترین موتور جستجوی **PPC** است ، که قبلاً به عنوان **GoTo** شناخته شده است .

**Teoma :** یک موتور جستجوی نسبتاً جدید است . درمقایسه با **oldies** مانند **altavista**

Yahoo یکی از اولین و دوست داشتنی ترین دایرکتوریهای وب است Yahoo .در حال حاضر (

2009 ) سایتی است که بیشترین بازدید کننده را در اینترنت دارد.

ریسمون نخستین موتور جستجوگر

ریسمون نخستین موتور جستجوگر مستقل منابع فارسی در شبکه جهانی وب است.

این جستجوگر همه وب سایتهای فهرست Link.ir را در دوره های زمانی یک ماهه می پیماید و محتویات و

مطالب آنها را نمایه سازی می کند و در بانک اطلاعاتی خود جهت ارائه خدمات جستجو به مراجعه کنندگان

نگهداری می نماید.

پروژه جستجوگر ریسمون از مهرماه 1383 در شرکت رادکام آغاز گردید و هم اکنون پس از بیش از یک سال

تلاش و پیگیری مستمر در مرحله راه اندازی آزمایشی قرار گرفته است.

در حال حاضر ریسمون به صورت آزمایشی راه اندازی شده است و ممکن است در ارائه خدمات خود دچار

مشکلاتی باشد. ما همه کوشش خود را در جهت رفع مشکلات آن و ارائه بهتر خدمات به کار خواهیم بست و در

این راه امیدواریم تا پیشنهادات و نظرات شما یاریگر ما باشد.

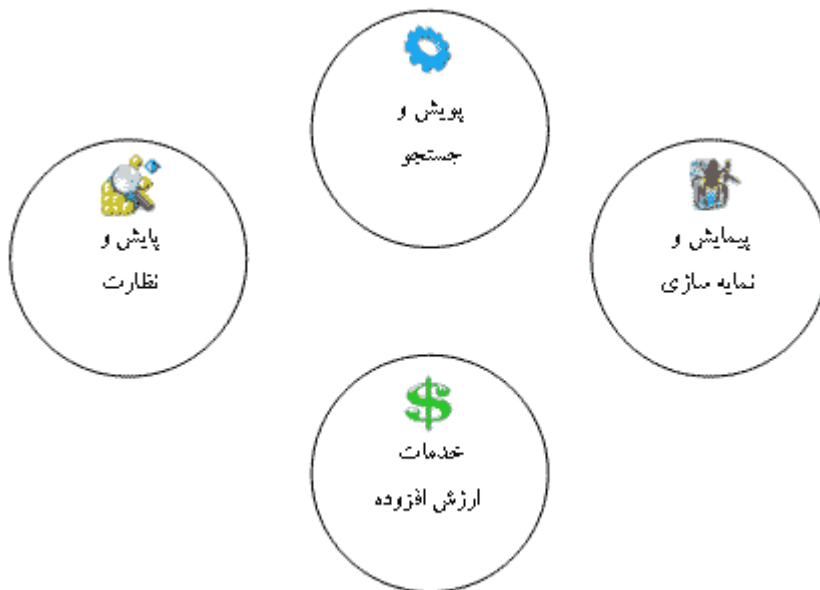
در اینجا لازم به ذکر است که پروژه ریسمون منحصر بوسیله شرکت رادکام به اجرا درآمده است و این شرکت در

طراحی، اجرا و راه اندازی آن به طور مستقل و بدون هیچ گونه پشتیبانی مادی و معنوی دیگران عمل نموده است

و در مورد ریسمون به هیچ نهاد دولتی و خصوصی و یا شخصی بدهکار نیست .

## معماری ریسمون

ریسمون دارای مؤلفه های اصلی یک موتور جستجو است. شکل 1 مؤلفه های یک موتور جستجوی مستقل نوعی، در بستر وب را نشان می دهد.



شکل 1: مؤلفه های موتور جستجوی مستقل

## پیمایش و نمایه سازی:

این مؤلفه سامانه ای است که با خزیدن در وب سایت ها، محتویات آنها را برای فرایند نمایه سازی و قراردادن در بانک اطلاعاتی، پالایش می کند. این سامانه معمولاً برای یافتن وب سایت ها، از فهرست مرجع خود استفاده می کند. موتورهای جستجوگر غول پیکری مانند Google، دیگر تنها به فهرست مرجع خود متکی نیستند(البته در ابتدا متکی بود) بلکه همه پیوندهای موجود در هر وب سایتی را به صورت زنجیره ای دنبال می کنند. اما در مورد



موتور های جستجوگر کوچک و یا منطقه ای با توجه به محدود بودن دامنه وب سایتها، فهرست مرجعی وجود دارد که سامانه پیمایش و نمایه سازی تنها به نمایه سازی وب سایتهای آن اقدام می نماید. در مورد ریسمن این فهرست مرجع، وب سایت [Link.ir](http://Link.ir) است که در بخش های بعدی به تفصیل در مورد آن صحبت خواهد شد.

#### پویش و جستجو:

این مؤلفه، بانک اطلاعاتی را که بوسیله مؤلفه پیمایش و نمایه سازی پر شده است را برای عبارات مورد نظر، جستجو می کند. در واقع هسته اصلی جستجوگر در این این مؤلفه نهفته است. در مورد این مؤلفه آنچه از همه مهمتر است زمان پاسخ آن است و باید الگوریتم های مناسب و نیز زیرساختهای سخت افزاری کارا، برای به حداقل رساندن زمان پاسخ این مؤلفه به کار گرفته شود.

#### پایش و نظارت:

این مؤلفه نقش نظارتی و تحلیلی بر خدمات موتور جستجوگر دارد. اطلاعات آماری که از این مؤلفه استخراج می گردد بسیار ارزشمند است و به عنوان مثال مشخص می کند که مردم بیشتر به چه چیزی علاقه دارند و به دنبال چه می گردند. علاوه بر این امکان ارائه اطلاعات آماری جغرافیایی نیز وجود دارد. تحلیل این اطلاعات آماری از دیدگاه های مختلف تجاری، فرهنگی و اجتماعی و حتی امنیت ملی بسیار مهم و قابل توجه است.

رزش افزوده:

با بالا رفتن مخاطبین وب سائتی که خدمات جستجو را به عنوان یکی از نیازهای اولیه کاربران وب، ارائه می دهد؛ توان اجرای طرحهای تجاری در این وب سایت بالا خواهد رفت. شاید تجربه Google در این عرصه مثال زدنی باشد. چرا که این وب سایت پس از جذب مخاطبین بیشمار از همه نقاط دنیا، با اجرای طرح تبلیغات در نتایج جستجو توانست به موفقیت تجاری چشم گیری دست یابد .

به علاوه این وب سایتها برای هرچه بالا رفتن تعداد مخاطبین باز هم به ارائه خدمات جدید اقدام می کنند تا کاربران را هرچه بیشتر به خدمات خود وابسته نگه دارند. باز هم در این زمینه شاید Google موفق ترین باشد، این وب سایت با عرضه خدماتی مانند Froogle، Orkut، GoogleEarth، Gmail و بسیاری خدمات دیگر، گوی سبقت را از سایر رقبای خود در جذب مخاطب ربوده است .

در ریسمون مؤلفه پیمایش و نمایه سازی با استناد به فهرست Link.ir به عنوان مرجع پیوند خود، در وب سایتها می -خزد و اطلاعات آنها را پس از استخراج و نمایه سازی در بانک اطلاعاتی خود ذخیره می کند. این فرایند در حال حاضر با توجه به زیرساختهای موجود برای نمایه سازی 2 میلیون صفحه، به مدت 10 روز به طول می انجامد. پس از پایان این فرایند بانک اطلاعاتی برای عملیات جستجو در اختیار مؤلفه پویش و جستجو قرار می گیرد و سامانه پیمایش و نمایه سازی، بانک اطلاعاتی قدیمی را دریافت و کار خود را دوباره آغاز می کند. به عبارت دیگر هر مؤلفه، یک بانک اطلاعاتی جداگانه برای خود دارد تا کار آنها با یکدیگر تداخل نداشته باشد. با این روش به حداکثر کارایی برای هر مؤلفه در تعامل آن با بانک اطلاعاتی دست می یابیم.

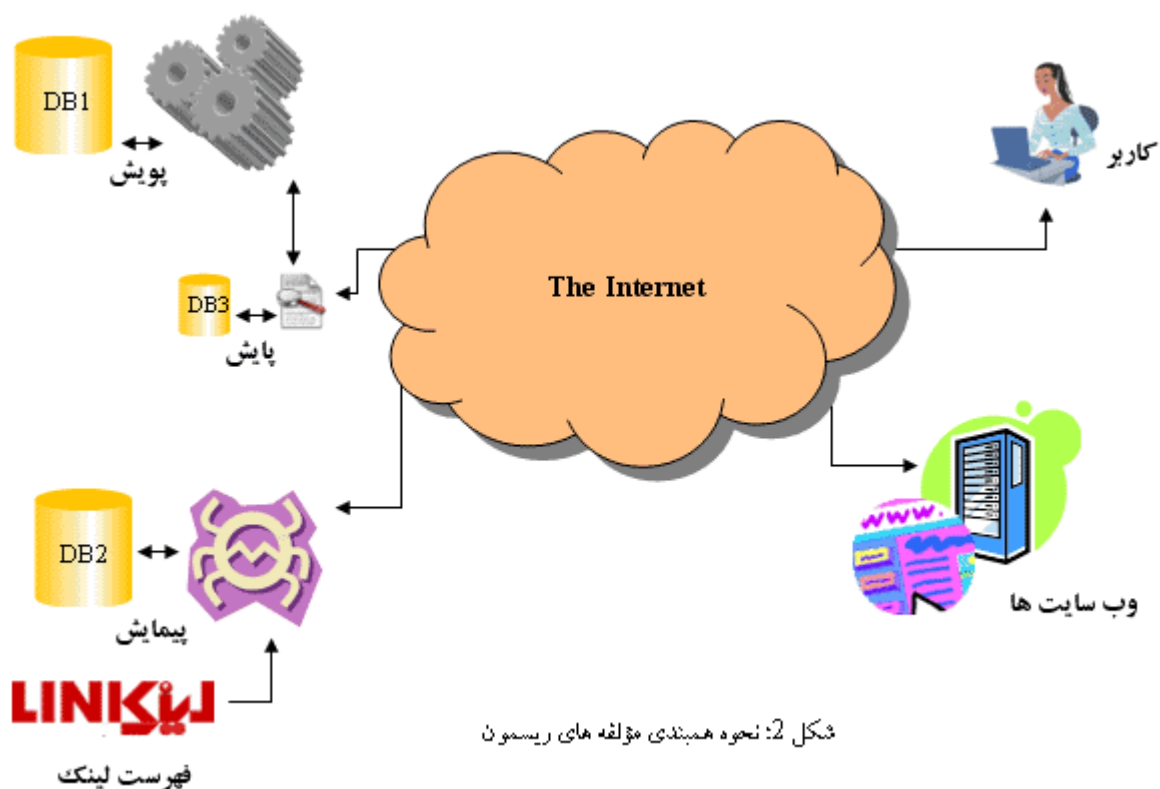
مؤلفه پایش و نظارت نیز به عنوان واسط بین کاربر و مؤلفه پویش و جستجو عمل می نماید؛ این مؤلفه عبارات مورد جستجو و سایر اطلاعات کاربر را جهت پردازش های بعدی ثبت می نماید.

مؤلفه خدمات ارزش افزوده نیز در ریسمون در نظر گرفته شده است. هم اکنون در وب سایت ریسمون جستجوی

اطلاعات تلفنی 118 کل کشور پیاده سازی شده است که از بانک اطلاعاتی کارگزارهای مرکز خدمات 118

مخابرات تهران استفاده می نماید. به علاوه یک سامانه تبلیغاتی نیز در حال پیاده سازی است که نسخه های اولیه

آن در فهرست Link.ir نصب و راه اندازی شده است.



قابلیت های عمومی ریسمون

قابلیت های عمومی ریسمون عبارتند از:

1. نمایه سازی اسناد و پرونده هایی با قالبهای Open Office, Text RTF, Microsoft Office, PDF, HTML با پشتیبانی کامل از زبان فارسی .
2. نمایه سازی محتویات فارسی بصورت جامع ، بطوریکه صفحاتی که با کاف و یای عربی تولید شده اند نیز به صورت فارسی نمایه سازی می شوند و در خروجی جستجو ظاهر می گردند .
3. نمایه سازی وب سایتهایی که با پروتکل SSL امن شده اند .
4. نمایه سازی اسناد و پرونده هایی که در FTP سایت قراردارند .
5. بهره مندی از Caching برای نگهداری اسناد و صفحات نمایه سازی شده .
6. زمانبندی نمایه سازی مجدد .
7. دسته بندی وب سایتهایی که نمایه سازی می شوند و امکان ارائه خدمات جستجو بصورت یک وب سرویس به وب سایتهای تابعه پورتالها دارد .
8. قابلیت اعتبارسنجی برای ورود به بخشهای اینترنتی که برای دسترسی به آنها به گذرواژه نیاز است .
9. پشتیبانی از stop words برای مشخص کردن کلمات و یا عباراتی که لازم نیست در نتیجه جستجو ظاهر شوند، مانند حروف اضافه "و"، "از"، "به"، ...
10. جستجوی منطقی بصورت ترکیب عطفی، فصلی و یا نفی از منطق بولی .

11. بهره مندی از جستجوی پیشرفته با قابلیت‌هایی نظیر محدود کردن جستجو به موضوع مورد نظر .
12. مشخص کردن کلمات یافت شده در نتیجه جستجو .
13. نمایش خلاصه صفحه و یا سند یافت شده در نتیجه جستجو .
14. اجرای چند نسخه از برنامه نمایه سازی و عملکرد همزمان آنها برای تسریع در امر نمایه سازی .
15. بهره مندی از روال رتبه بندی نتایج جستجو به طوریکه صفحات و اسناد با ارتباط بیشتر، رتبه بالاتری در خروجی جستجو دارند .
16. نمایش تعداد کل نتایج یافت شده .
17. محاسبه و نمایش زمانی که صرف جستجو شده است .

### Zoom Search قابلیت منحصر به فرد ریسمون

از جمله قابلیت‌های منحصر به فرد جستجوگر ریسمون ارائه خدمات جستجو با امکان محدود کردن دامنه موضوع جستجو است. این قابلیت، امکان یافتن مطالب بدون تعویض عبارت مورد جستجو و صرفاً با تحدید موضوع، جستجو را برای کاربر فراهم می کند.

هم اکنون در ریسمون امکان تحدید نتایج جستجو بر حسب موضوعات موجود در فهرست [Link.ir](http://Link.ir) تعبیه شده است و این امر به پالایش نتایج جستجو بوسیله کاربر متناسب با موضوع خود و یافتن صفحات و مطالب مورد نظر او کمک شایانی می کند.

به علاوه با این روش حتی می توان در ریسمون جستجوهای تخصصی ایجاد کرد. هم اکنون بخش جستجوی "شعر و ادب" در ریسمون به عنوان یک خروجی تخصصی جستجو عمل می نماید .

به عبارت دیگر این امکان نوعی Zoom در نتایج جستجو است و از امکانات منحصر به فرد ریسمون نسبت به سایر رقبای خارجی آن می باشد.

معرفی فهرست Link.ir به عنوان مرجع پیوندهای ریسمون

ریسمون از فهرست Link.ir به عنوان مرجع پیوند خود استفاده می کند. علاوه بر این فهرست لینک امکان مرور موضوعی وب سایتهای ایرانی را نیز فراهم می کند. امکانی که از جهت ویژگیهای خود، تنها یک فهرست، می تواند در اختیار قرار دهد، چرا که هدف جستجوگر ها یافتن و دسترسی به محتویات وب سایت ها است و اگر کاربر بخواهد فهرست وب سایتهایی که در یک موضوع خاص فعالیت می کنند را بیابد، هیچ ابزاری به جز یک فهرست این امکان را در اختیار او قرار نمی دهد .

در حال حاضر در فهرست Link.ir افرادی به عنوان ویراستار سرفصل های اصلی، فعالیت می کنند؛ که مسوولیت دسته بندی و درج پیوند وب سایتهای سرفصل زیر مجموعه خود را دارند. کار دسته بندی وب سایتها و وارد کردن آنها در شاخه های مرتبط، کاری است که شاید در نگاه اول ساده به نظر برسد، ولی بسیار وقت گیر است و دقت خاص خود را می طلبد چرا که فردی که این کار را انجام می دهد باید آشنایی لازم را با نوع فعالیت وب سایت داشته باشد تا بتواند آن را در دسته مرتبط با آن ثبت کند.

ویراستار هر سرفصل پس از بررسی و مطالعه وب سایت و حصول اطمینان از حداقل های شرایط ثبت در فهرست Link.ir، بهترین جایگاه را در میان شاخه زیر مجموعه خود انتخاب و وب سایت را به همراه توضیحاتی درباره وب سایت و نیز اطلاعات تماس با آن وب سایت، در آن شاخه درج می نماید. پس از این فرایند یک نامه الکترونیکی به Webmaster وب سایت ثبت شده، ارسال می گردد و او را از این امر مطلع می کند تا اگر نظر خاصی درباره شاخه، یا توضیحات وب سایت ثبت شده داشت، ویراستار را مطلع گرداند.