
Scientific Search Engines

By
Atefeh Davoudian

Zanjan University of Medical Sciences
Summer 1401

تعریف موتور جستجو

■ موتورهای جستجو درحقیقت، ماشینهای پاسخ دهی هستند. موتور جستجو یک سیستم نرم افزاری است که وظیفه اصلی آن پیدا کردن محتوا اعم از محتوای متنی، صوتی، عکس، ویدئو و هر فایل دیگری در شبکه جهانی وب است؛ محتوایی که بیشترین مطابقت را با کوئری شما داشته باشد.

■ یک موتور جستجو از دو بخش اصلی تشکیل شده است:

○ پایگاه داده اطلاعات

○ الگوریتم ها

موتورهای جستجو چگونه کار می کنند؟

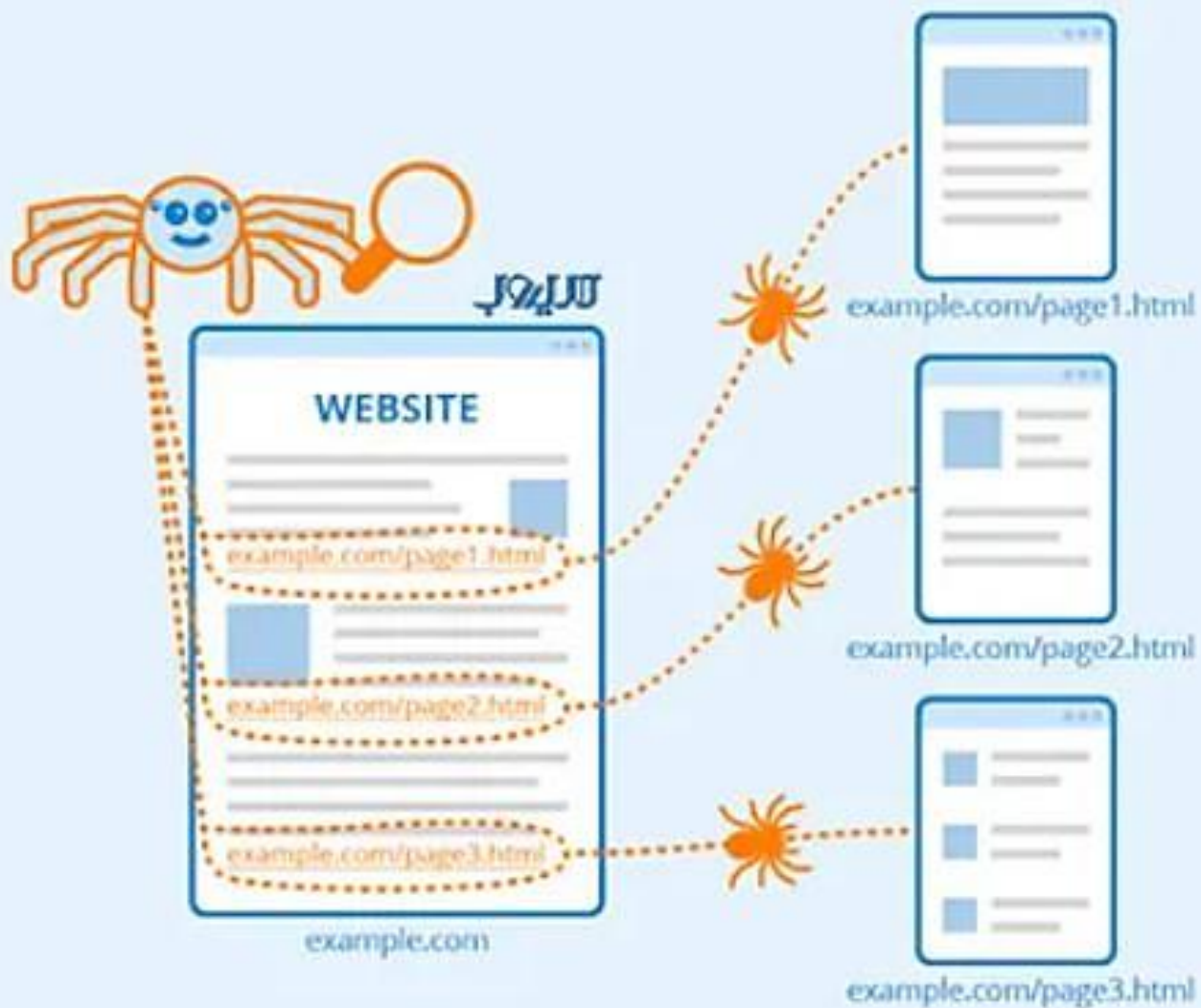
- فرآیند کشف، دسته بندی و رتبه بندی میلیون ها وبسایتی که در فضای اینترنت وجود دارند، از طریق ۳ اقدام بنیادی انجام می پذیرد:

Crawling (۱)

Indexing (۲)

Ranking (۳)

- با استفاده از این اقدامات، موتورهای جستجو محتواهای تازه منتشر شده را کشف میکنند، آنها را در پایگاه داده خود ذخیره میکنند و در نهایت برای استفاده کاربران طبقه بندی میکنند



موتورهای جستجو چگونه کار می‌کنند؟

کراال کردن (Crawling)

ایندکس کردن (Indexing)

رتبه‌بندی کردن (Ranking)

Crawling

- موتورهای جستجو ربات هایی در اختیار دارند که به خزنده (**Crawler**) و عنکبوت نیز معروفند. آنها ربات های خود را برای ارزیابی محتوای وبسایت های تازه ایجاد شده و یا محتوایی که به تازگی تغییر کرده باشند ارسال می کنند. وظیفه این ربات ها این است که صفحات وب را به دقت بررسی کنند تا دریابند که هر کدام در چه زمینه های بوجود آمده اند و چه موضوعی دارند. سپس، با دنبال کردن لینک های موجود در صفحات، به صفحات دیگر رفته و آن صفحات را نیز کراال می کنند.

Indexing

- بعد از اینکه یک وبسایت کرال شد، موتورهای جستجو باید تصمیم بگیرند که چطور آن اطلاعات را طبقه بندی کنند. فرآیند ایندکس کردن به ذخیره سازی صفحات کرال شده در دیتابیس های موتورهای جستجو گفته می شود. هر موتور جستجو دیتابیس مخصوص به خود را دارد.

Ranking

■ وقتی یک کاربر جستجویی انجام می دهد، موتورهای جستجو به سراغ دیتابیس خود رفته تا اطلاعات مرتبطی را از بین تمام آن صفحاتی که قبلا ایندکس کرده بودند، ارائه کنند. اما این اطلاعات به چه ترتیبی باید ارائه شوند؟ تمام صفحات یافت شده، براساس یک سیستم رتبه دهی برای کاربران فهرست می شوند و به هیچ وجه ترتیبی تصادفی ندارند. موتورهای جستجو فرآیند رتبه بندی را با در نظر گرفتن فاکتورهای مختلف و براساس کیفیت و مرتبط بودن با کوئری کاربر انجام می دهند.

چگونه موتورهای جستجو اهمیت یک مطلب را تعیین می کنند؟

- به طور معمول یک موتور جستجو مقدار اهمیت یک مورد را بر اساس محبوبیت آن تعیین می کند. طبق این فرمول هر چقدر یک سایت یا صفحه بیشتر محبوب و پر بازدید باشد، از اهمیت و ارزش بالاتری در بخش اطلاعات درون خود برخوردار است.
- این محبوبیت یا مرتبط بودن به صورت دستی تعیین نمی شوند. این مقادیر توسط هوش مصنوعی موتورهای جستجو، محاسبات دقیق ریاضی و الگوریتم ها معین می شوند. این الگوریتم ها معمولاً شامل هزاران بخش مختلف هستند.

مهمترین فاکتورهای اثرگذار بر روی الگوریتم های موتور جستجو

- **مرتبط بودن عنوان و محتوا:** اینکه تا چه حد عنوان و محتوای صفحه به کوئری کاربران مرتبط است.
- **نوع محتوا:** اگر کاربران به دنبال عکس هستند، نتایج شامل عکسها خواهند بود نه متن.
- **کیفیت محتوا:** محتوا باید کامل، مفید، آموزنده و بیطرفانه باشد.
- **کیفیت وبسایت:** کیفیت کلی یک وبسایت اهمیت بسیاری دارد و حتی المقدور باید طبق استانداردهای آن موتور جستجو باشد.
- **سرعت صفحه:** صفحاتی که سریعتر بارگذاری می شوند، شانس بیشتری در مقایسه با صفحات کند دارند.

مهمترین فاکتورهای اثرگذار بر روی الگوریتم های موتور جستجو

- **تاریخ انتشار:** برای کوئری های مرتبط با اخبار یا سایر مواردی که بروز بودن محتوا اهمیت خاصی دارد. مثل سایت های هواشناسی یا نرخ ارز، تاریخ انتشار محتوا نیز در رتبه بندی سایت دخیل است.
- **محبوبیت صفحه:** این مورد به میزان ترافیک یک سایت بستگی ندارد؛ بلکه به درک دیگر وبسایت ها از آن صفحه خاص برمی گردد. صفحه ای که بک لینک های متعددی از وبسایت های دیگر دارد، از صفحات دیگری که هیچ لینکی نداشته یا لینک کمتری دارند محبوبتر به حساب می آید.
- **زبان صفحه:** صفحه به زبان کاربران خاص خودش نوشته شود.

THE TOP LIST OF ACADEMIC SEARCH ENGINES



1. Google Scholar

Google Scholar is the clear number one when it comes to academic search engines. It's the power of Google searches applied to research papers and patents. It not only let's you find research papers for all academic disciplines for free, but also often provides links to full text PDF file.

- Coverage: approx. 200 million articles
- Abstracts: only a snippet of the abstract is available
- Related articles: ✓
- References: ✓
- Cited by: ✓
- Links to full text: ✓
- Export formats: APA, MLA, Chicago, Harvard, Vancouver, RIS, BibTeX

https://scholar.google.com/



scholar.google.com



My profile



My library



Google Scholar



☒ Articles ☐ Case law

Stand on the shoulders of giants

2. BASE

BASE is hosted at Bielefeld University in Germany and that's where its name stems from (Bielefeld Academic Search Engine).

- Coverage: approx. 136 million articles (contains duplicates)
- Abstracts: ✓
- Related articles: ✗
- References: ✗
- Cited by: ✗
- Links to full text: ✓
- Export formats: RIS, BibTeX

https://www.base-search.net/



base-search.net



Login

English



Basic search

Advanced search

Browsing

Search history



BASE

Bielefeld Academic Search Engine

Search 311,977,910 documents from 10,152 content providers



3. RefSeek

RefSeek searches more than one billion documents from academic and organizational websites. Its clean interface makes it especially easy to use for students and new researchers.

- Coverage: no detailed statistics available, approx. 1 billion documents
- Abstracts: only snippets of the article are available
- Related articles: ✗
- References: ✗
- Cited by: ✗
- Links to full text: ✓
- Export formats: not available

https://www.refseek.com/

← → ↻ 🏠 🔒 refseek.com

🔗 ☆ 🧩 🖨️ 👤 ⋮

☰ Directory 🔍 Search

refseek*

Web | Documents

Search



Learn about: [Telephones](#), [Neptune](#)

Browse the [Reference Site Directory](#)



4. Baidu Scholar

Although **Baidu** Scholar's interface is in **Chinese** it's index contains research papers in English as well as Chinese.

- Coverage: no detailed statistics available, approx. 100 million articles
- Abstracts: only snippets of the abstract are available
- Related articles: ✓
- References: ✓
- Cited by: ✗
- Links to full text: ✓
- Export formats: APA, MLA, RIS, BibTeX

https://xueshu.baidu.com/

← → ↻ 🏠 xueshu.baidu.com

🔍 🔗 ☆ 🧩 🖨️ Ⓐ ⋮

网页 新闻 贴吧 知道 图片 视频 地图 文库 | 百度首页 登录 注册



高级搜索

百度一下

推荐

订阅主题词，兴趣内容首页展示 >

期刊 磁悬浮导轨上的力学实验误差分析 [全文](#)

磁悬浮实验套装在一定程度上普及了磁悬浮技术的应用,但是由于仪器设计的自身限制,使得实验误差比较大,文章在分析了引起实验误差的基础上,既对实验指导提出了可行性意见,又在仪器研发...

李爽 周晓红 | 科技视界 | 2020

🗣️ 0次被引

❤️ 喜欢

★ 收藏

站内功能

首单免费



论文查重



学术分析



开题分析



Jurn

Jurn is a free online search tool for the finding and downloading of **free full-text** scholarly works. Jurn is powered by a **Google Custom Search Engine (CSE)** and is run without any adverts.

- Coverage: 800 ecology titles, Directory of 3,000 arts & humanities journals
- Abstracts: only snippets of the abstract are available
- Related articles: ✗
- References: ✗
- Cited by: ✗
- Links to website: ✓
- Export formats: APA, MLA, RIS, BibTeX ✗

https://www.jurn.link/#gsc.tab=0

← → ↻ 🏠 🔒 jurn.link/#gsc.tab=0

🔗 ☆ 🧩 🖨️ 👤 A



Search millions of free academic articles, chapters and theses.

| Arts | Humanities | Business | Law | Nature | Science | Medical |

ENHANCED BY Google



Index updated: 10th March 2022.

| [How to use JURN](#) | [FAQ](#) | [Blog](#) | [A-Z of 800 ecology titles in JURN](#) |
| [Directory of 3,000 arts & humanities journals in JURN, of 5,054 indexed](#) |

New! [Search all the world's academic repositories, full-text and records alike.](#)

Virtual Learning Resources Center

The **Virtual LRC** is a meta-search engine that includes in its results information gleaned from many of the best research portals and university and public library Internet subject guides. It's a completely free resource, is the creation of Dr. Michael Bell, former state chair of the Texas Association of School Librarians.

- Coverage: collection of more than 10,000 indexed webpages for all subject areas
- Abstracts: only snippets of the abstract are available
- Related articles: X
- References: ✓
- Cited by: X
- Links to website: ✓
- Export formats: APA, MLA, RIS, BibTeX X

https://www.virtuallrc.com/



virtuallrc.com



Welcome to VLRC: Sunday, Sept. 04, 2022 10:28:25 AM



*Virtual Learning
Resources Center* **VirtualLRC.com**

Enhanced By Google



Directory