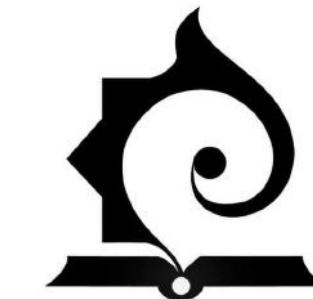


موتورهای جستوجو

محمدجواد رخشانی

mrakhshani4d@gmail.com



دانشکده فنی و مهندسی شهید باهنر شیراز - بهار ۱۴۰۲

جویشگر^[۱]

نرم افزاری برای جستجو در محیطهای مانند اینترنت که در آنها حجم داده‌ها بسیار زیاد است.

۱. «جویشگر» همارز «Search Engine»؛ منبع: گروه واژه‌گزینی و زیر نظر حسن حبیبی، دفتر دوم، فرهنگ واژه‌های مصوب فرهنگستان، تهران: انتشارات فرهنگستان زبان و ادب فارسی، شابک ۰-۳۷-۷۵۳۱-۹۶۴

کلمه‌های کلیدی

Beethoven

ai GitHub Copilot
ChatGPT OpenAI
Microsoft

روز مادر

کرونا ویروس

شیراز قیمت ارز

اخبار

آدرس محل ذخیره

<https://aparat.com>

<https://openai.com>

<https://microsoft.com>

<https://bahonarshiraz.tvu.ac.ir>

<https://news.yahoo.com>

<https://shirazmetro.ir>

<https://samad.app>

کلمه‌های کلیدی

• آدرس محل ذخیره

OpenAI

1. <https://openai.com>
2. <https://open.ai>
3. <https://www.youtube.com>
4. <https://learn.microsoft.com> > azure
5. <https://openaimaster.com>

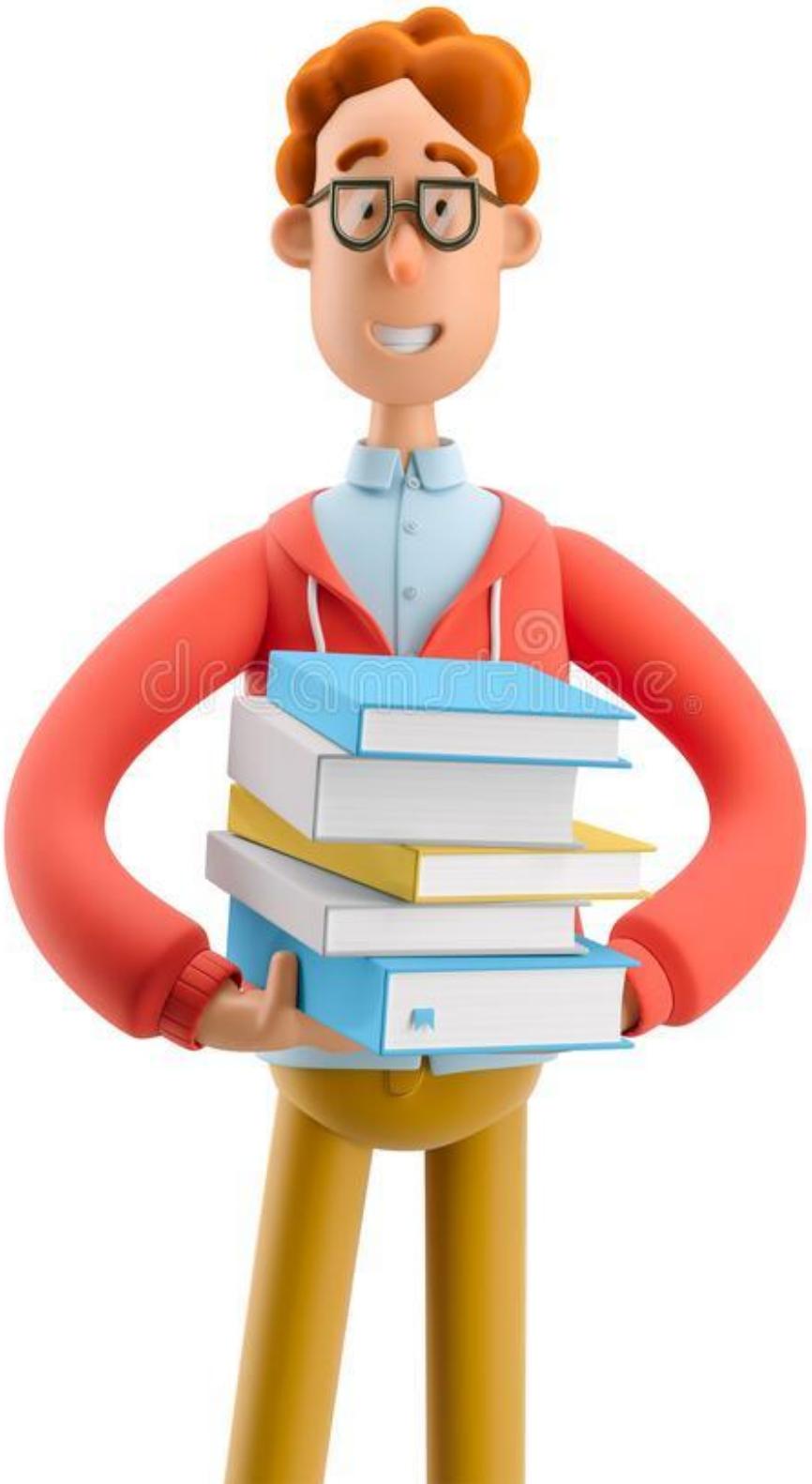
کلمه‌های کلیدی

• آدرس محل ذخیره

شیراز

1. <https://fa.wikipedia.org>
2. <https://www.shiraz.ir>
3. <https://shirazu.ac.ir>
4. <https://shiraz.iau.ir>
5. <https://www.sums.ac.ir>

انواع موتورهای جستوجو؟



انواع موتورهای جستوجو؟

۱. موتورهای جستوجو انحصاری و عمومی؛
۲. موتورهای جستوجو آزمایشی و تجاري؛
۳. موتورهای جستوجو مبتنی بر درونکاو، فهرست و مختلط.

موتورهای جستوجو انحصاری

به طور خاص برای جستوجو در یک وبسایت خاص طراحی و تولید می‌شوند.

☰ YouTube DE

politecnico di torino X Q :

Accedi

Home Filtri Informazioni su questi risultati i

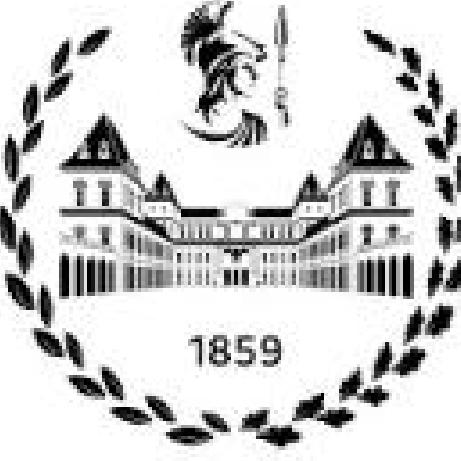
Shorts

Iscrizioni Iscriviti

Raccolta

Cronologia

Le novità di Politecnico di Torino



Politecnico di Torino
@PolitecnicodiTorinoChannel • 8420 iscritti
Il Politecnico di Torino è una Research university con circa 30.000 studenti. In questo canale

ELEZIONI STUDENTESCHE Biennio 2023-2025

22-23-24-25 maggio 2023 in modalità telematica



1:20:58

Elezioni studentesche 2023-25 | Presentazione liste
104 visualizzazioni • 4 giorni fa

Politecnico di Torino

15 maggio 2023.

Nuovo

CUG - Comitato Unico di Garanzia |

تصوير ١. موتور جوستجوی داخلی وBSITE YouTube

دیوان 

کتاب دیوان حافظ انتشارات پیام عدالت 

کتاب دیوان کامل فروغ فرخزاد انتشارات نگاه آشنا 

کتاب دیوان حافظ اثر خواجه شمس الدین محمد... 

دیوان 

در دسته **کتاب شعر و ادبیات**

کتاب دیوان 

در دسته **کتاب**

کتاب دیوان فروغ فرخزاد 

در دسته **کتاب فلسفه و روانشناسی**

دیوان 

دیوان حافظ 

تصویر ۲. موتور جوستجوی داخلی وبسایت دیجی‌کالا

موتورهای جستجو عمومی

محتوای صفحه‌های تمام وب را کاوش و در پایگاهداده‌هایش ذخیره می‌کند.

Chi siamo Google Store

Gmail Immagini



Accedi

Google



Cerca con Google

Mi sento fortunato

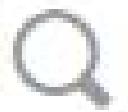
Italia

تصویر ۳. موتور جوستجوی Google



DuckDuckGo

Cerca nel web senza venire spiato



**Ti scoccia essere spiato online?
Possiamo aiutarti.**

Proteggi la tua privacy online gratuitamente con un solo download:

Ricerca Privata

Blocco Dei Sistemi Di Tracciamento

Navigazione Criptata

تصویر ©. موتور جوستجوی DuckDuckGo

ذره بین



جستجو کنید

گلچین خبر ورزش سرگرمی خانواده هنر علم و فناوری الالعنه سلامت



درباره ما

تماس با ما

دانلود اپلیکیشن

فروختن

بارجورد

تصویر ۵. موتور جوستجوی ذره بین

انواع موتورهای جستوجو؟

۱. موتورهای جستوجو انحصاری و عمومی؛
۲. موتورهای جستوجو آزمایشی و تجاری؛
۳. موتورهای جستوجو مبتنی بر درونکاو، فهرست و مختلط.

موتورهای جستوجو آزمایشی

موتورهای جستوجوی تحقیقاتی که برای آزمایش نوع خاصی از الگوریتم‌ها ساخته شده‌اند.

انواع موتورهای جستوجو؟

• موتورهای جستوجو آزمایشی

دانشگاهها یا موسسه‌های تحقیقاتی

انواع موتورهای جستوجو؟

• موتورهای جستوجو آزمایشی

توسعه به عنوان موتور جستوجو تجاری

موتورهای جستوجو تجاری

برای کاوش و جستوجو در کل پنهان اینترنت به کار می‌روند.

انواع موتورهای جستوجو؟

• موتورهای جستوجو تجاری

جلب رضایت کاربران

انواع موتورهای جستوجو؟

• موتورهای جستوجو تجاری

درآمد های تبلیغاتی

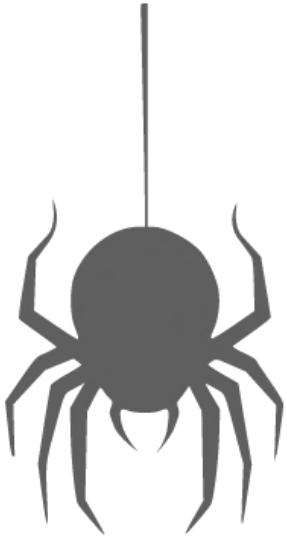
انواع موتورهای جستوجو؟

• موتورهای جستوجو تجاری

پشتیبانی مالی

انواع موتورهای جستوجو؟

۱. موتورهای جستوجو انحصاری و عمومی؛
۲. موتورهای جستوجو آزمایشی و تجاری؛
۳. موتورهای جستوجو مبتنی بر درونکاو، فهرست و مختلط.



موتورهای جستجو مبتنی بر درون کاو^[۱]

موتورهای جستجوی هوشمند و خودمختار در فرآیند دریافت، دخیره، ردیابی و شاخص‌دهی.

انواع موتورهای جستوجو؟

• موتورهای جستوجو مبتنی بر درون کاو

هوشمند و خودمختار^[۱]

1. Autonomous

انواع موتورهای جستوجو؟

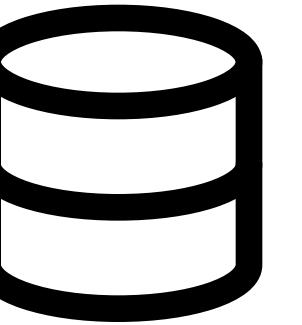
- موتورهای جستوجو مبتنی بر درون کاو

دريافت سند 

انواع موتورهای جستجو؟

- موتورهای جستجو مبتنی بر درون کاو

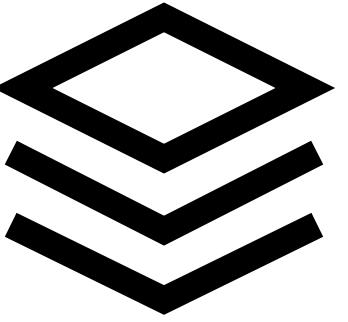
ذخیره سند



انواع موتورهای جستوجو؟

- موتورهای جستوجو مبتنی بر درون کاو

ردہبندی سند



انواع موتورهای جستوجو؟

• موتورهای جستوجو مبتنی بر درون کاو

شاخص دهنده سند

موتورهای جستجو مبتنی بر فهرست^[۱]

محتوای ثبت شده از سندها در ساختارهایی به نام «فهرستها» سازماندهی و ذخیره شده‌اند.

انواع موتورهای جستوجو؟

• موتورهای جستوجو مبتنی بر فهرست

دخلت مستقیم صاحب سندها

انواع موتورهای جستوجو؟

• موتورهای جستوجو مبتنی بر فهرست

ثبت، نظارت و به روزرسانی

انواع موتورهای جستوجو؟

• موتورهای جستوجو مبتنی بر فهرست

تغییرات محتویات درونی سندها



Search by Keyword

BROWSE CATEGORIES BROWSE BY REGION PRICING

GET LISTED TODAY

Best of the Web is the oldest, editor curated web directory online. Whether you're looking for a plumber, a lawyer that specializes in a specific legal field, or need someone to design your website, Best of the Web has you covered.

Research and review our verified listings by category to help you choose the business right for you.



Automotive

Find the best car dealerships, trusted neighborhood auto mechanics, or top car washes near you. Search the best of the best in the automotive industry here.

Automotive Repair

Towing Companies

Auto Body Repair Shops

Auto Dealers



Business

Browse top business websites across the B2B sector, as well as top shopping and services sites. Whatever your industry, the most important business websites are listed here.

Business Classifieds

Business Directories

Business Resources

Business to Business

تصویر ۶. فهرست The Best of the Web

انواع موتورهای جستوجو؟

• موتورهای جستوجو مبتنی بر فهرست

مدیر فهرست‌ها

[۱]

مоторهای جستوجو مختلط

تلفیقی از یک سیستم فهرست غنی و یک مotor جستوجو برای جستوجو در فهرست.

پک یا چند سیستم فهرست

+

پک موتور جستوجوی قدرتمند



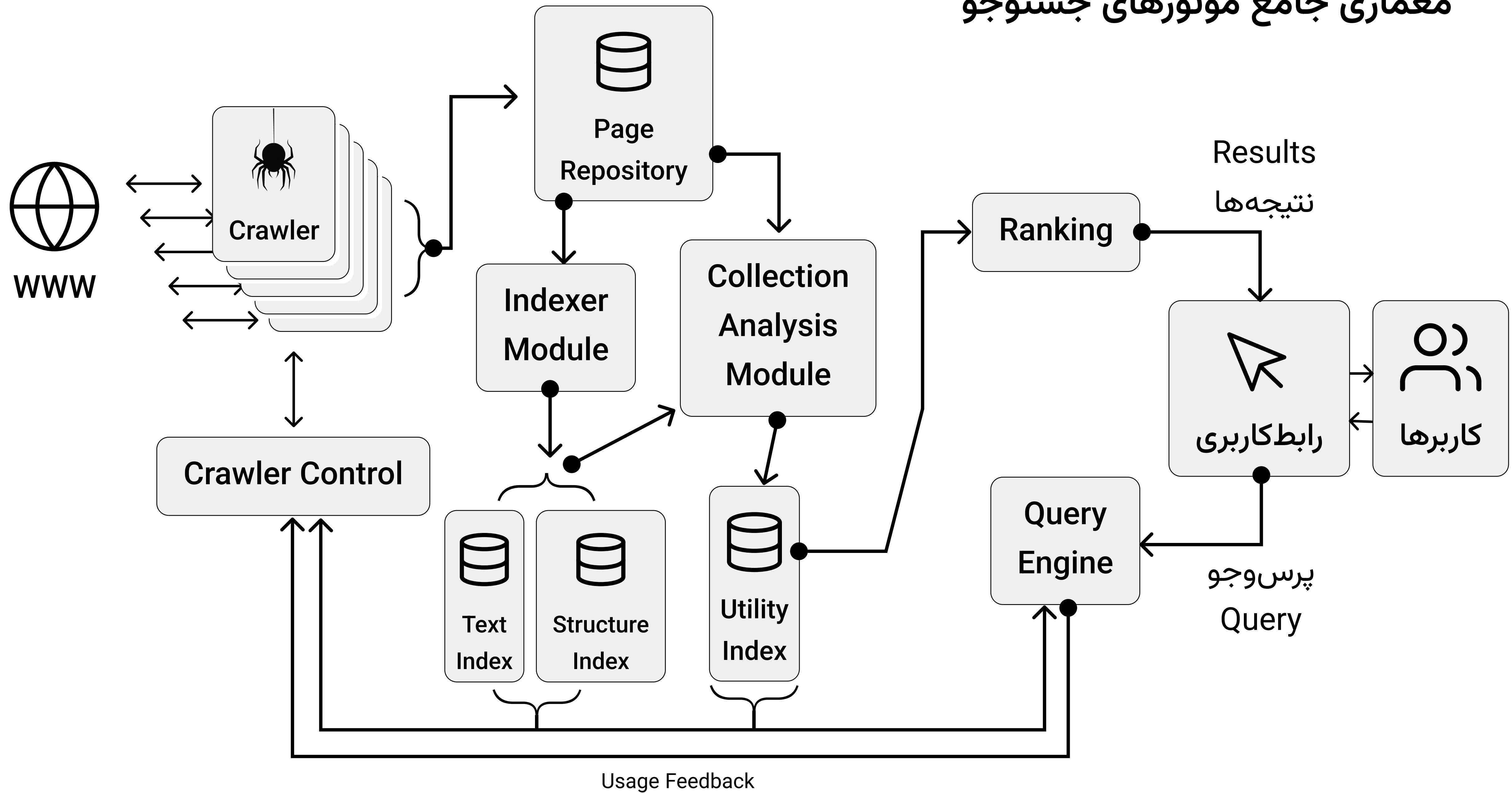
انواع موتورهای جستوجو؟

۱. موتورهای جستوجو انحصاری و عمومی،
۲. موتورهای جستوجو آزمایشی و تجاری،
۳. موتورهای جستوجو مبتنی بر درون کاو، فهرست و مختلط.

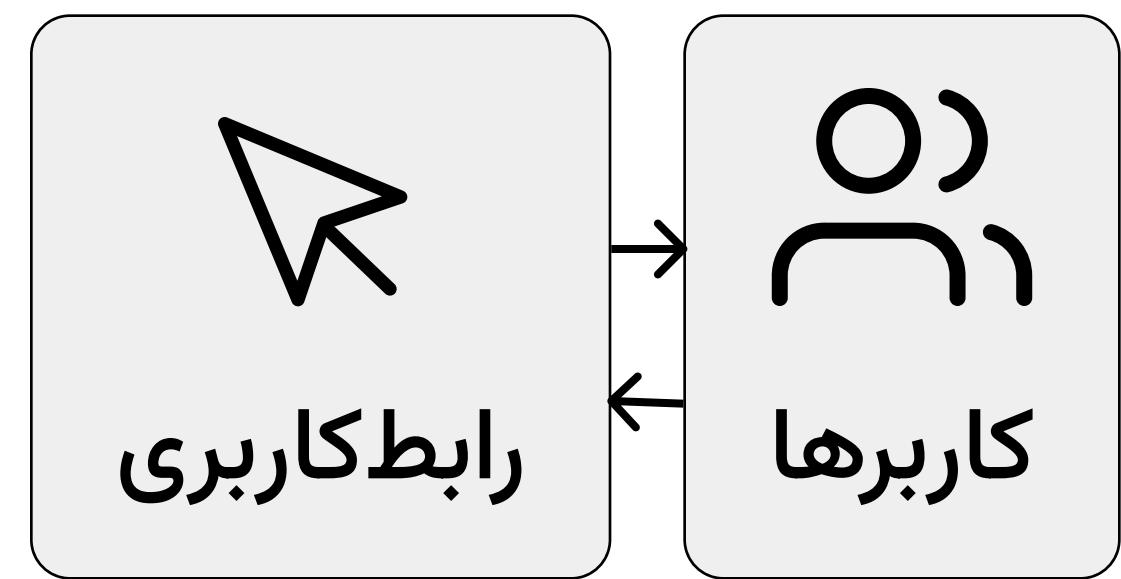
معماری جامع موتورهای جستجو

مоторهای جستجو به دلایل تجاری و امنیتی در تلاش‌اند تا الگوریتم‌های درونی‌شان را پنهان کنند.

معماری جامع موتورهای جستجو



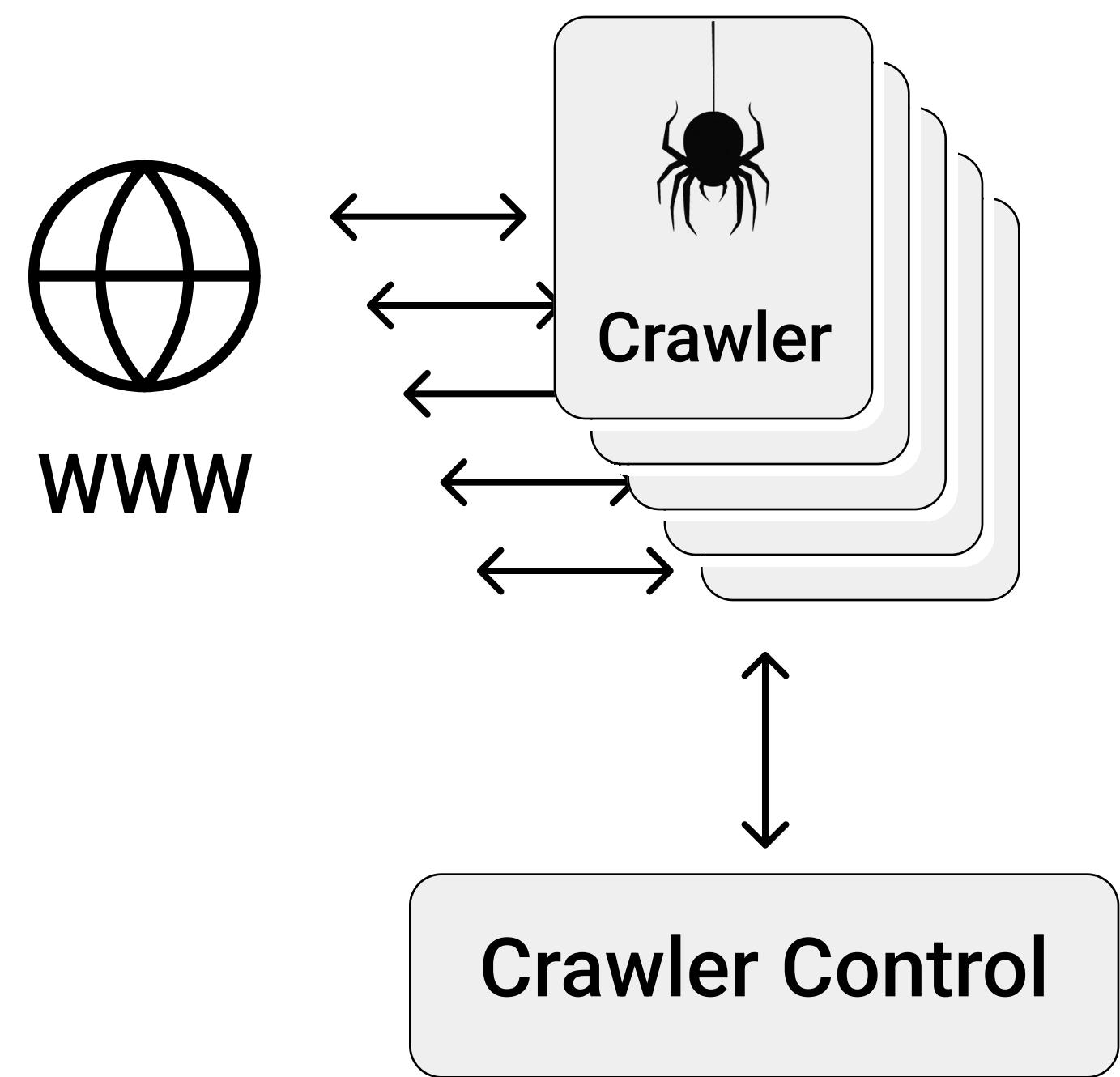
معماری جامع موتورهای جستجو



معماری جامع موتورهای جستجو

Query
Engine

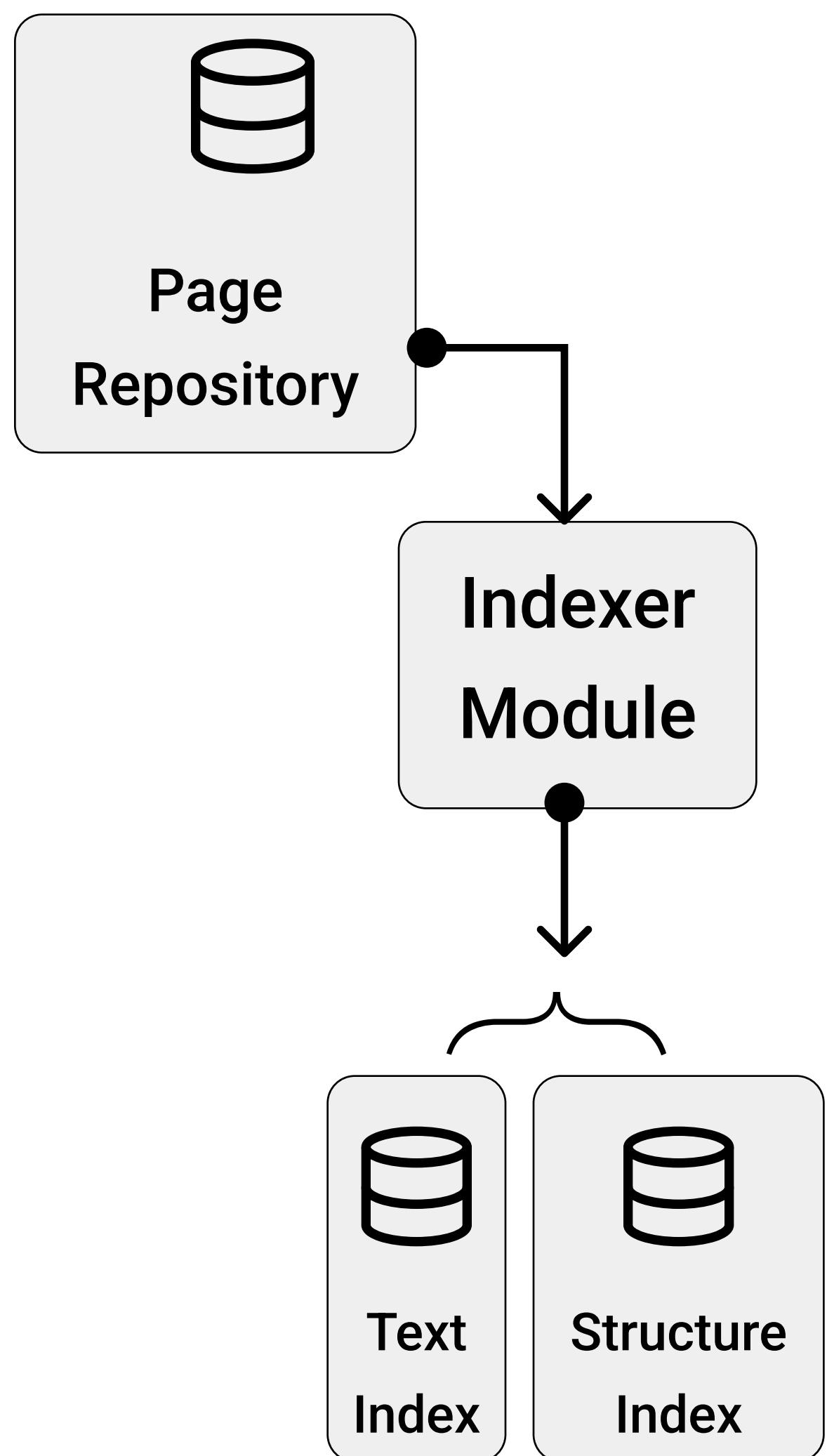
معماری جامع موتورهای جستجو



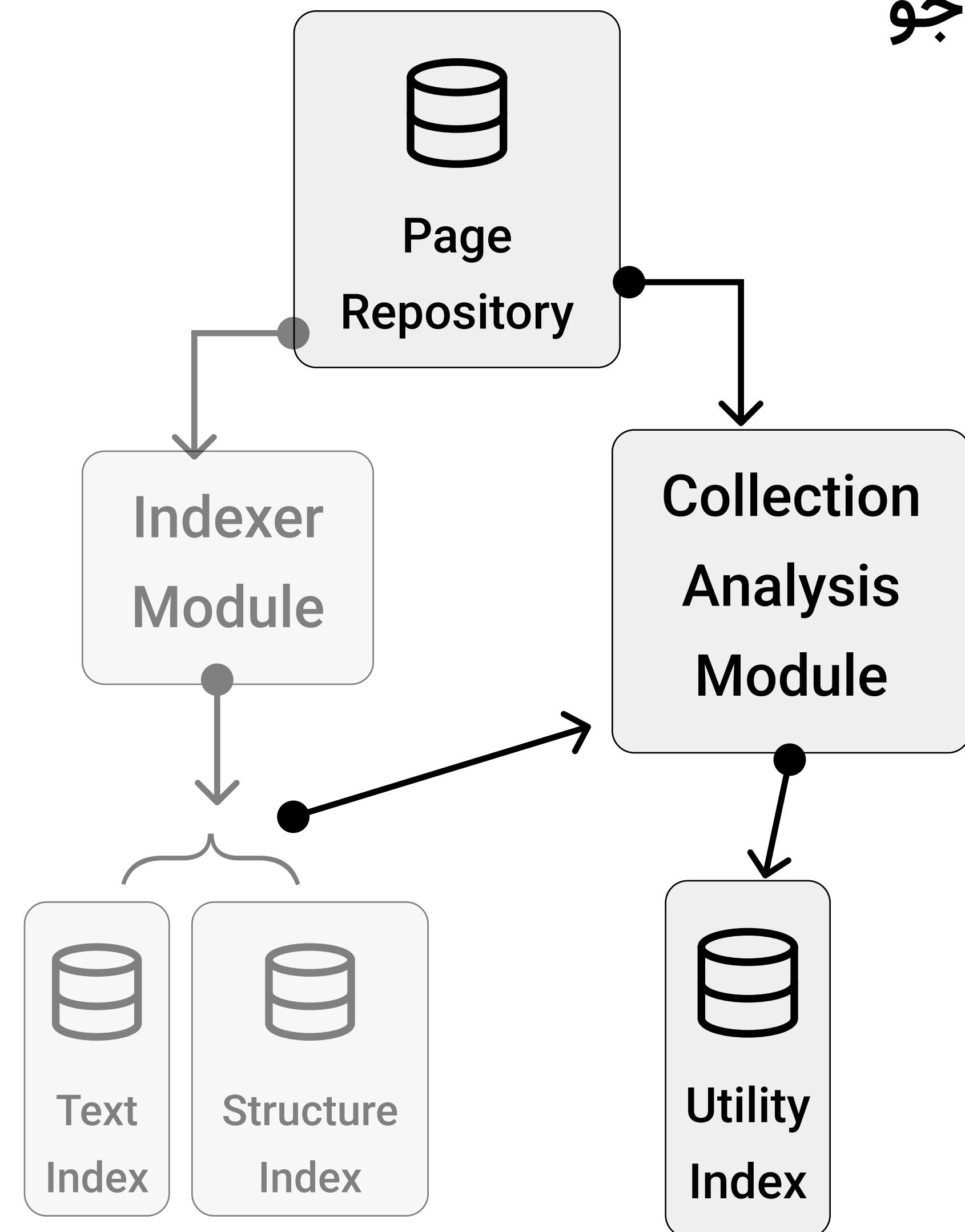
معماری جامع موتورهای جستجو



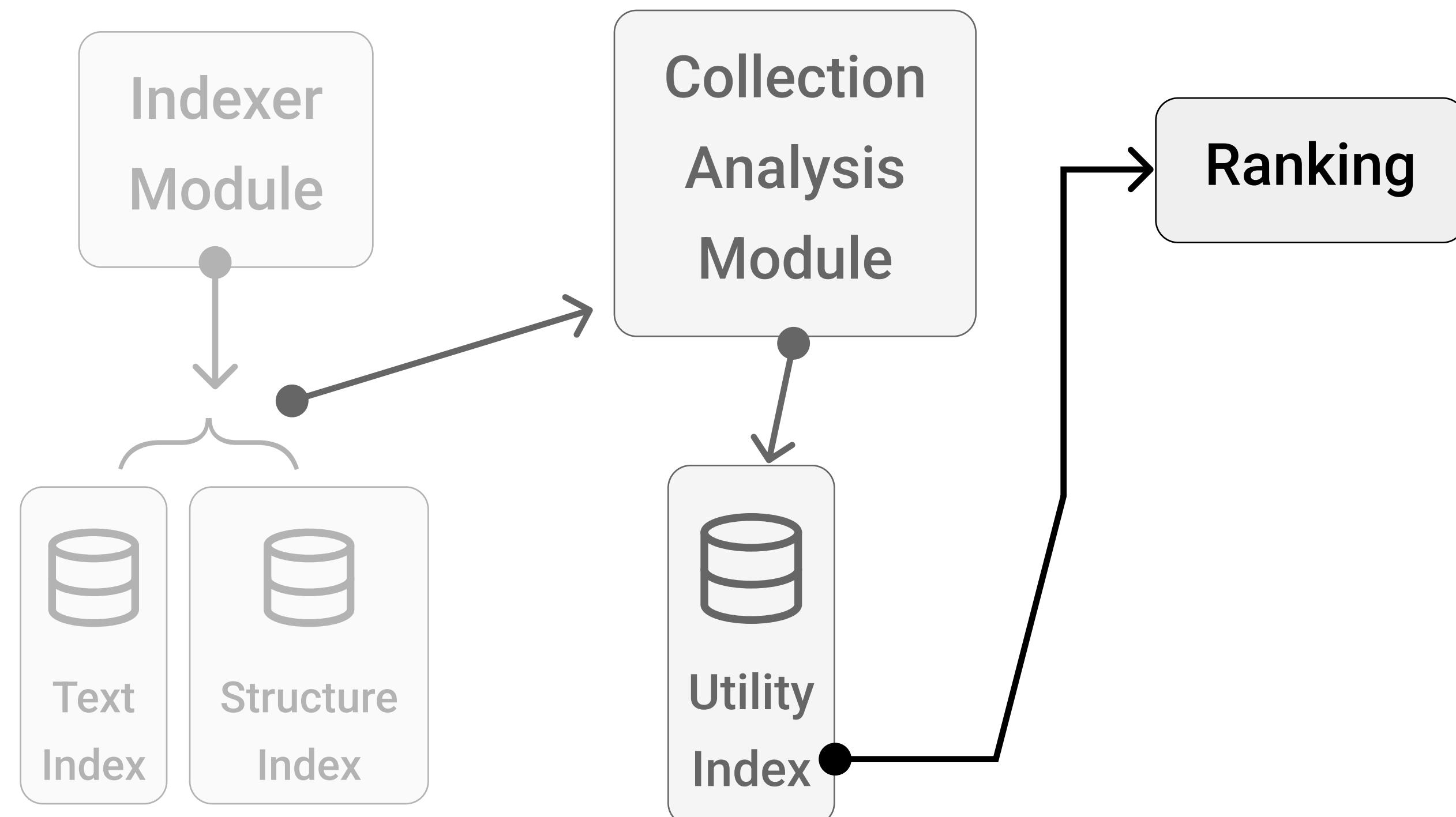
معماری جامع موتورهای جستجو



معماری جامع موتورهای جستجو



معماری جامع موتورهای جستجو



انتخاب صفحات برای درونکاوی

یکی از مهم‌ترین چالش‌های «درونکاو وب» انتخاب صفحه برای دریافت و ذخیرهسازی است.

معیار انتخاب صفحات برای درونکاوی

۱. مبتنی بر گرایش‌های کاربر؛
۲. مبتنی بر شهرت صفحه‌ها؛
۳. مبتنی بر محل قرار گرفتن صفحات.

معیار انتخاب صفحات برای درون کاوی

• مبتنی بر گرایش‌های کاربر

مجموعه بزرگی از پرس‌و‌جوهای کاربرها

معیار انتخاب صفحات برای درون کاوی

• مبتنی بر گرایش‌های کاربر

کلمه‌های کلیدی

معیار انتخاب صفحات برای درونکاوی

• مبتنی بر گرایش‌های کاربر

فراوانی کلمه‌کلیدی مورد نظر در یک صفحه

معیار انتخاب صفحات برای درونکاوی

۱. مبتنی بر ~~گرایش‌های کاربر~~؛
۲. مبتنی بر شهرت صفحه‌ها؛
۳. مبتنی بر محل قرار گرفتن صفحات.

معیار انتخاب صفحات برای درونکاوی

• مبتنی بر شهرت صفحه‌ها

«شهرت و اعتبار» برای سنجش یک صفحه
اما چطور؟

معیار انتخاب صفحات برای درونکاوی

• مبتنی بر شهرت صفحه‌ها

Backlink Count

معیار انتخاب صفحات برای درونکاوی

۱. مبتنی بر گرایش‌های کاربر؛
۲. مبتنی بر شهرت صفحه‌ها؛
۳. مبتنی بر محل قرار گرفتن صفحات.

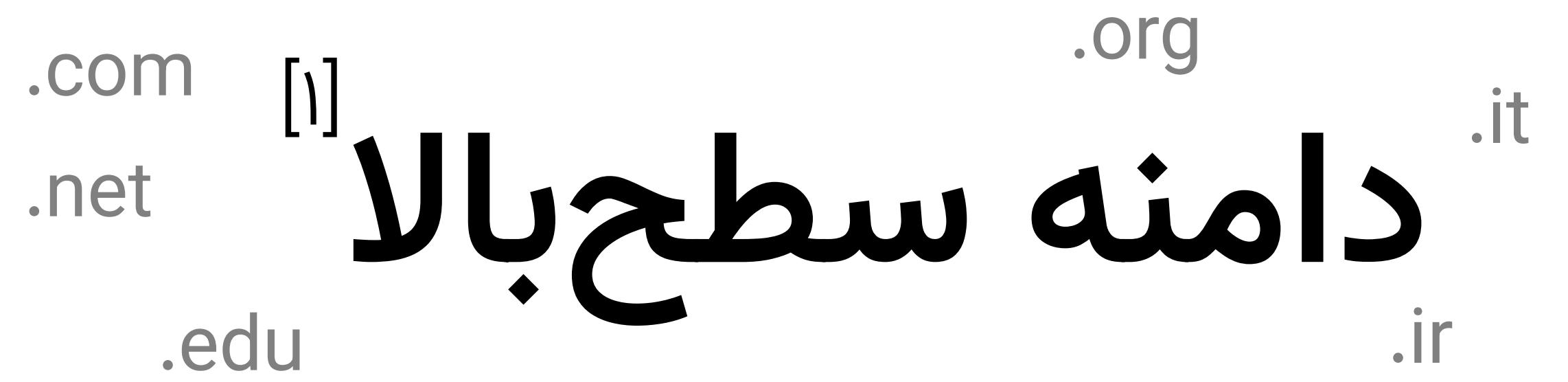
معیار انتخاب صفحات برای درونکاوی

• مبتنی بر محل قرارگرفتن صفحات

نام صفحه

معیار انتخاب صفحات برای درون کاوی

- مبتنی بر محل قرارگرفتن صفحات



1. Top-level domain

معیار انتخاب صفحات برای درونکاوی

• مبتنی بر محل قرارگرفتن صفحات

میزان فاصله از «صفحه ریشه»^[۱]

روش‌های به روزرسانی و سرکشی درونکاو

۱. به روزرسانی یکنواخت؛
۲. به روزرسانی متناسب با تغییر.

روش‌های به روزرسانی و سرکشی درون کاو

• به روزرسانی یکنواخت

سرکشی در زمان مشخص شده

روش‌های به روزرسانی و سرکشی درون‌کاو

• به روزرسانی یکنواخت

فارغ از این‌که صفحه تغییر کرده است با خیر

روش‌های به روزرسانی و سرکشی درونکاو

۱. به روزرسانی یکنواخت؟
۲. به روزرسانی متناسب با تغییر.

روش‌های به روزرسانی و سرکشی درون‌کاو

• به روزرسانی متناسب با تغییر

سرکشی بر اساس تناوب

روش‌های به روزرسانی و سرکشی درون کاو

• به روزرسانی متناسب با تغییر

تناوب سرکشی < تناوب تغییر

اطلاعات قدیمی‌تر در موتور جستجو

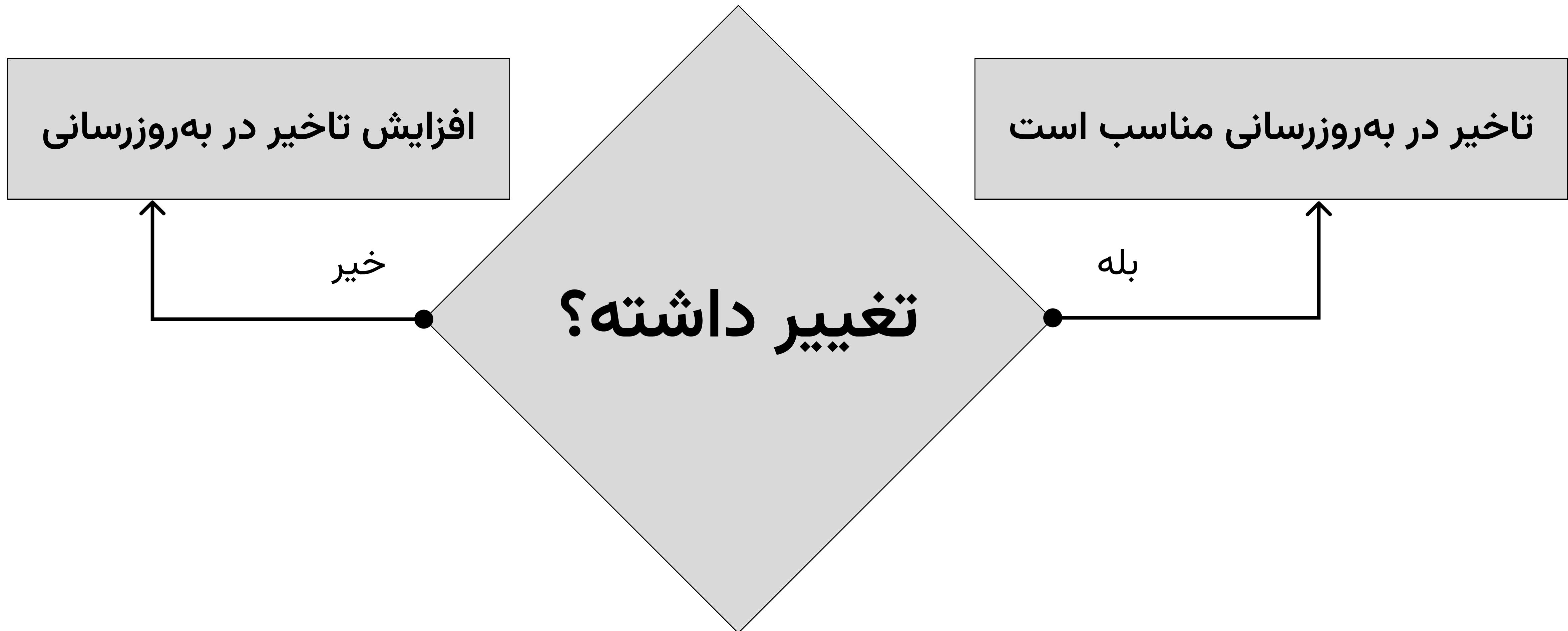
روش‌های به روزرسانی و سرکشی درون کاو

• به روزرسانی متناسب با تغییر

تناوب سرکشی اولیه

روش‌های به روزرسانی و سرکشی درون‌کاو

- به روزرسانی متناسب با تغییر



Meta Description Tag

برچسب‌های توصیفی (MDT)

```
<meta name="robots" content="noindex,nofollow">
```

برچسب‌های توصیفی (MDT)

<meta name="robots" content="noindex,nofollow">

name="robots"

برچسب‌های توصیفی (MDT)

<meta name="robots" content="noindex,nofollow">

content="noindex,nofollow"

فَایل robots.txt

فایل robots.txt

```
User-agent:  
Disallow:
```

<https://www.example.com/robots.txt>

فایل robots.txt

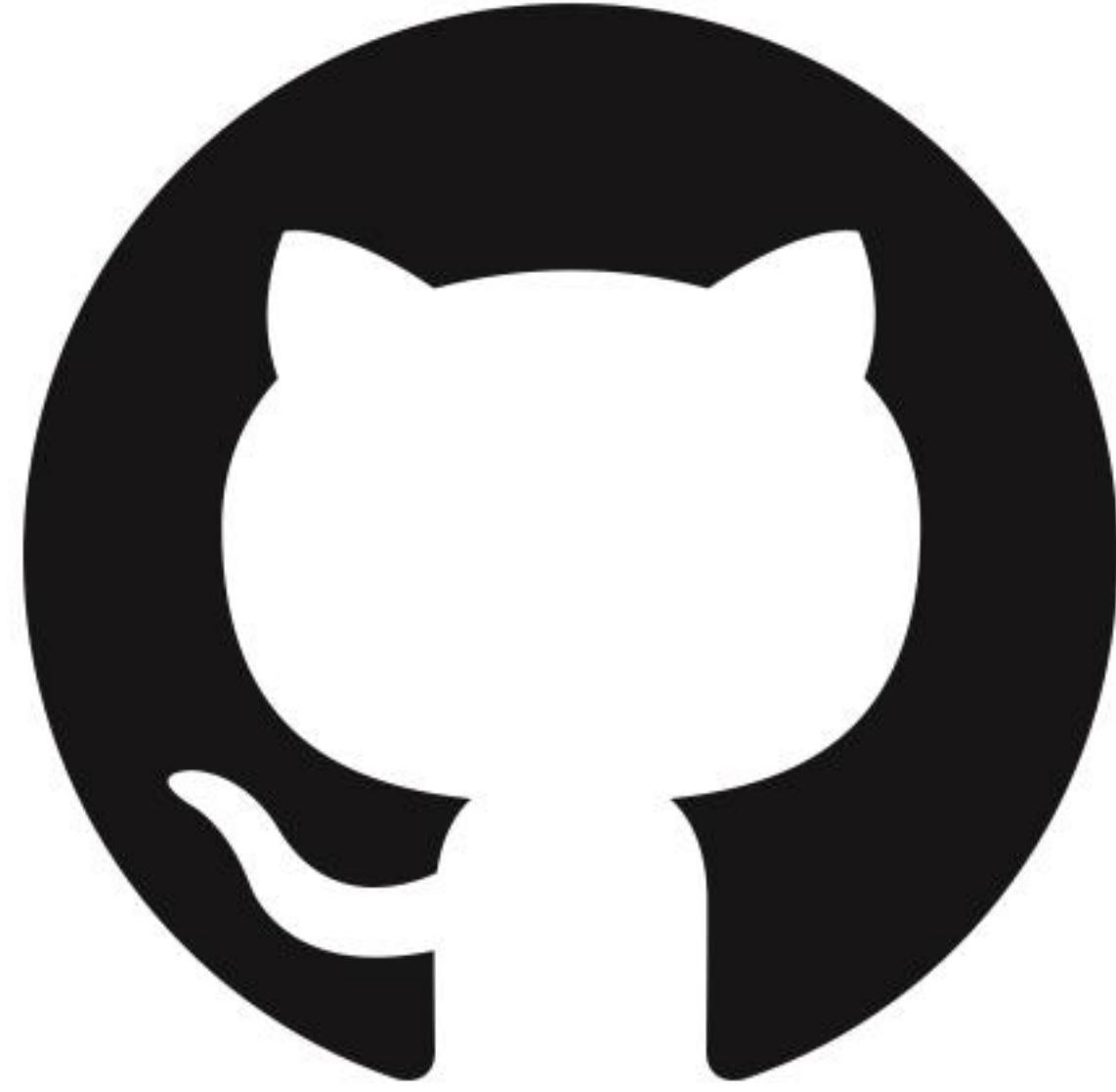
```
User-agent: *
Disallow: /
```

<https://www.example.com/robots.txt>

فایل robots.txt

```
User-agent: googlebot          # all Google services  
Disallow: /private/             # disallow this directory  
  
User-agent: googlebot-news      # only the news service  
Disallow: /                      # disallow everything  
  
User-agent: *                    # any robot  
Disallow: /something/           # disallow this directory
```

<https://www.example.com/robots.txt>



<https://github.com/loreys/mlscraper>

متشکرم!



<https://github.com/mohammadjr7>

mrakhshani4d@gmail.com