

برای آشنایی مقدماتی با لوسین می‌توان به مقالات فارسی ذیل مراجعه کرد:

[راهنمای گام به گام Lucene در c#](#)

[کتابخانه جستجوی فارسی با Lucene.net](#)

[آشنایی با کتابخانه Lucene.NET - قسمت اول](#)

[آشنایی با کتابخانه Lucene.NET - قسمت دوم](#)

[معرفی کتابخانه‌ی مشهور و قدرتمند Lucene.net -- قسمت اول](#)

به صورت خلاصه اگر نیاز به جستجوی سریع و پیشرفته‌ای بر روی حجم عظیمی از اطلاعات دارید، روش متداول `select * from table where field like something` توصیه نمی‌شود. بسیار کند است؛ مصرف CPU بالایی دارد. از ایندکس استفاده نمی‌کند. راه حل توصیه شده جهت برخورد با این نوع مسایل استفاده از `full text search` است. نگارش کامل SQL Server حاوی یک موتور FTS [توکار هست](#). اگر از بانک اطلاعاتی خاصی استفاده می‌کنید که دارای موتور FTS نیست یا ... FTS مخصوص SQL Server به درد کار شما نمی‌خورد یا نیاز به سفارشی سازی دارد (مثلا امکان تعریف stop words فارسی (کلماتی مانند به، از، تا و امثال آن))، از موتور FTS جانبی دیگری به نام [لوسین](#) نیز می‌توان استفاده کرد.

در کنار این‌ها ابزاری برای آنالیز و کوئری گرفتن از فایل‌های ایندکس تهیه شده توسط لوسین نیز وجود دارد به نام [Luke](#). برای نمونه اگر بانک اطلاعاتی سایت جاری را با لوسین به نحو متداولی ایندکس کنیم، در صفحه اول این برنامه، `top ranking terms` به شکل زیر ظاهر می‌شود:

Top ranking terms. (Right-click for more options)			
No	Rank ▼	Field	Text
1	871	Body	br
2	870	Body	div
3	865	Body	style
4	864	Body	align
5	856	Body	dir
6	852	Body	right
7	835	Body	rtl
8	810	Body	در
9	800	Body	از
10	797	Body	و
11	785	Body	text
12	778	Body	src
13	776	Body	به
14	774	Body	http
15	771	Body	href
16	767	Body	img
17	743	Body	1
18	730	Body	class
19	726	Body	های
20	711	Body	post
21	708	Body	width
22	707	Body	را
23	706	Body	height

Show top terms >>

Number of top terms:

100

Hint: use Shift-Click to select ranges, or Ctrl-Click to select multiple fields (or unselect all).

Tokens marked in red indicate decoding errors, likely due to a mismatched decoder.

در اینجا چون متون تهیه شده از نوع HTML هستند، تگ br در آن‌ها زیاد است و یا یک سری حروف و کلمات فارسی هم در صدر قرار دارند که بهتر است از لیست ایندکس حذف شوند. برای اینکار تنها کافی است یک hash table را به نحو زیر تعریف و به StandardAnalyzer لوسین ارسال کنیم:

```
var stopWords = new Hashtable();
stopWords.Add("br","br");
// ...
var analyzer = new StandardAnalyzer(Version.LUCENE_29, stopWords);
```

یا آقای عرب عامری برای حروف و کلمات فارسی که نباید ایندکس شوند، یک لیست نسبتاً جامع را [در اینجا](#) تهیه کرده‌اند. اینبار اگر stop words یاد شده را اعمال و مجدداً ایندکس‌ها را تهیه کنیم به خروجی بهتری خواهیم رسید. در کل حداقل از این لحاظ، لوسین نسبت به FTS توکار SQL Server مناسب‌تر به نظر می‌رسد.

نظرات خوانندگان

نویسنده: شهرز جعفری
تاریخ: ۱۳۹۱/۰۴/۲۶ ۰:۸

سلام

یه سری از این نمونه‌ها از کد زیر استفاده شده

```
////////// Begin the new section
QueryParser oParser = new QueryParser("Body", new StandardAnalyzer());
string sTitle = "", sWriterID = "", finalQuery = "";

sTitle = " AND (Title:" + titleTerm + ")";

sWriterID = " AND (WriterID:" + writerID + ")";

finalQuery = "(" + bodyTerm + sTitle + sWriterID + ")";
hits = searcher.Search(oParser.Parse(finalQuery));
////////// End of the new section
```

که به نظرم حرفه ای نیست.
در ضمن اگه میشه چند تا سایت برای استفاده از کتابخانه در Entity framework معرفی کنید.

نویسنده: وحید نصیری
تاریخ: ۱۳۹۱/۰۴/۲۶ ۰:۴۷

کاری به EF نداره. به شکل یک سیستم مستقل بهش نگاه کنید. رکوردها از Db دریافت و به شکل document به لوسین اضافه خواهند شد. در همین حین index هم تشکیل می‌شود.
کوئری‌های آن دقیقاً به همین شکلی هست که در بالا اومده و زبان آن [SQL نیست](#).
البته پروژه LINQ به آن هم وجود دارد: ([^](#))

نویسنده: رضا.ب
تاریخ: ۱۳۹۱/۰۴/۲۷ ۱۱:۵۴

برای full text search در زبان فارسی این 150 کلمه که آقا عرب‌عامری زحمت کشیدن اصلاً کفایت نمی‌کند. زبان فارسی و دیگه‌اش به این نحو اصلاً مناسب ایندکس کردن نیست و بیشتر نیاز به توسعه دارد.
شاید اگر بشه با ابزارهای توسعه‌ای که روی زبان فارسی کار شده - مثل ویراسباز [^](#) - خروجی بگیریم برای ایندکس کردن و در نتیجه تحلیل اونا به وسیله همین کتابخونه Luke یا Solr یا Nutch - که در وبلاگ آقای زبردست معرفی شده [^](#) - آسونتره.

نویسنده: وحید نصیری
تاریخ: ۱۳۹۱/۰۴/۲۷ ۱۳:۵

کار آقای عرب‌عامری فراتر است از 150 کلمه ذکر شده. آنالیز فتحه، کسره، ی فارسی و عربی و غیره هم در آن لحاظ شده.

نویسنده: رضا.ب
تاریخ: ۱۳۹۱/۰۴/۲۸ ۰:۰

جستجو که در سایت همکنون هست، از همین کتابخانه استفاده می‌کنه یا توسعه داده شده؟
یه سوال شاید بی‌ربط. در آنالیز نتایج شاخص‌های لوسین ضربی غیر از تعداد دفعات تکرار شاخص مد نظر هست؟

نویسنده: وحید نصیری
تاریخ: ۱۳۹۱/۰۴/۲۸ ۰:۲۶

- از کتابخانه اصلی lucene.net استفاده شده.
- توضیحات مفصلش رو می‌تونید [اینجا](#) مطالعه کنید.

نویسنده: مهدی پایروند
تاریخ: ۱۳:۳۶ ۱۳۹۱/۰۷/۱۰

مهندس پیشنهاد اجرای Luke در ویندوز 7 چی؟
این آدرس [لینک](#) کلی کامپوننت برای دانلود گذاشته

نویسنده: وحید نصیری
تاریخ: ۱۳:۳۹ ۱۳۹۱/۰۷/۱۰

- فایل‌های jar نیاز به موتور اجرایی جاوا (JRE) دارند.
- luke [نسخه دات نتی](#) هم دارد.

نویسنده: مهدی پایروند
تاریخ: ۱۵:۲۵ ۱۳۹۱/۰۷/۱۲

واقعا کتابخانه سریعی هستش برای کاری که من داشتم نزدیک به 16 هزار رکود 10 فیلدی رو توی لوسین ایندکس کرد که برای دریافت نتایجش زمان خیلی کمتری رو گرفت.

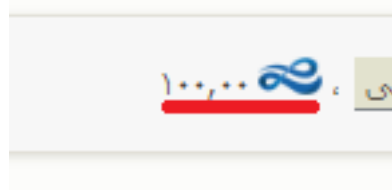
نویسنده: مهدی پایروند
تاریخ: ۱۵:۲۹ ۱۳۹۱/۰۷/۱۲

ممون برای من خیلی مورد استفاده بوده

نویسنده: مهدی پایروند
تاریخ: ۱۵:۳۷ ۱۳۹۱/۰۷/۱۲

در مورد این تصویر کمی توضیح میدین که به کدوم بخش لوسین مربوطه ممنون

از ایندکس استفاده نمی‌کند.



نویسنده: وحید نصیری
تاریخ: ۱۷:۱۱ ۱۳۹۱/۰۷/۱۲

این عدد درصد نزدیک بودن جواب به جستجوی انجام شده است (رتبه جستجوی لوسین). فرمول محاسبه آن به صورت زیر است:

```
var hits = searcher.Search(query, 10).ScoreDocs;  
var scoreNorm = 100.0f / hits.GetMaxScore();  
foreach (var scoreDoc in hits)  
{  
    var resultScore = scoreNorm * scoreDoc.score;  
}
```

```
}
```

نویسنده: وحید نصیری
تاریخ: ۱۷:۲۶ ۱۳۹۱/۰۷/۱۲

بله. اینقدر باکیفیت است که RavenDB برای سیستم جستجوی خودش از لوسین به صورت توکار استفاده می‌کند.

نویسنده: مهدی پایروند
تاریخ: ۲۳:۵ ۱۳۹۱/۰۷/۱۳

آیا برای صفحه بندی معادلی برای skip و take وجود دارد، با توجه به جستجویی که من انجام دادم ([+](#) و [+](#) و [+](#) و [+](#) و [+](#) و [+](#)) نمونه‌ای دال بر وجود یک همچنین امکانی ندیدم.

نویسنده: وحید نصیری
تاریخ: ۲۳:۳۵ ۱۳۹۱/۰۷/۱۳

- سؤال شما مرتبط با بحث Luke نیست.
- و ... نداره. هر بار باید جستجو کنید. بعد روی نتیجه حاصل [صفحه بندی رو](#) انجام بدید. یا می‌تونید نتیجه جستجوی خاص انجام شده رو کش کنید تا سربار جستجوی مجدد حذف شود. (هرچند اینقدر سریع است که نیازی به کش نیست)

نویسنده: مهدی پایروند
تاریخ: ۱۹:۴۴ ۱۳۹۱/۰۷/۱۴

بابت پاسختون ممنونم