

نحوه اضافه کردن قابلیت غلط گیر املائی شبیه به جستجوی گوگل توسط لوسین

عنوان:

وحید نصیری

نویسنده:

۲۱:۱۰ ۱۳۹۱/۰۹/۰۶

تاریخ:

www.dotnettips.info

آدرس:

Lucene.NET

گروه‌ها:

پیشنیاز:

[چگونه با استفاده از لوسین مطالب را ایندکس کنیم؟](#)

مقدمه

اگر به جستجوی سایت دقت کرده باشید، قابلیتی تحت عنوان پیشنهاد «عبارات مشابه» به آن اضافه شده است:



این مورد بر اساس ماژول غلط یاب املائی لوسین تهیه شده و بسیار شبیه به "did you mean" جستجوی گوگل است. در ادامه به نحوه پیاده سازی آن خواهیم پرداخت.

کتابخانه‌های مورد نیاز

علاوه بر [کتابخانه لوسین](#)، نیاز به دریافت پروژه [Contrib](#) آن نیز می‌باشد تا بتوان از اسمبلی Lucene.Net.ContribSpellChecker.dll موجود در آن استفاده کرد.

نحوه کار با غلط یاب املائی لوسین

خلاصه کار با غلط یاب املائی لوسین همین چند سطر ذیل است:

```
var indexReader = IndexReader.Open(FSDirectory.Open(indexPath), readOnly: true);  
  
// Create the SpellChecker  
var spellChecker = new SpellChecker.Net.Search.Spell.SpellChecker(FSDirectory.Open(indexPath +  
"\\Spell"));  
  
// Create SpellChecker Index  
spellChecker.ClearIndex();  
spellChecker.IndexDictionary(new LuceneDictionary(indexReader, "Title"));  
spellChecker.IndexDictionary(new LuceneDictionary(indexReader, "Body"));
```

```
//Suggest Similar Words  
var results = spellChecker.SuggestSimilar(term, number, null, null, true);
```

کار بر اساس یک ایندکس از پیش موجود لوسین شروع می‌شود. در اینجا فرض شده است که این ایندکس در پوشه indexPath قرار دارد.

در ادامه شیء spellChecker را آغاز خواهیم کرد. بهتر است پوشه تولید فایل‌های آن با پوشه ایندکس اصلی یکسان نباشد. اگر یکسان در نظر گرفته شود، تمام مداخل جدید به ایندکس موجود اضافه خواهند شد که می‌تواند سرعت جستجوی معمولی را کاهش دهد.

سپس کار تهیه ایندکس جدید غلط یاب املایی، شروع خواهد شد. متد spellChecker.ClearIndex، اطلاعات موجود در ایندکسی قدیمی را حذف کرده و سپس spellChecker.IndexDictionary، فیلدهایی را که نیاز داریم در تهیه غلط یاب املایی حضور داشته باشند، مشخص می‌کند.

همانطور که ملاحظه می‌کنید ایندکس جدید تهیه شده، بر اساس بانک اطلاعاتی واژه‌های موجود در ایندکس اصلی برنامه که توسط indexReader معرفی شده، تهیه می‌شود. برای نمونه در تصویر ابتدای مطلب جاری، واژه‌های پیشنهادی، واژه‌هایی هستند که بیشتر یکبار تایپ شده و در بانک اطلاعاتی برنامه موجود بوده‌اند.

و در آخر برای استفاده از امکانات تهیه شده، تنها کافی است متد spellChecker.SuggestSimilar را فراخوانی کنیم (در زمانیکه جستجوی اصلی سایت نتیجه‌ای را ارائه نداده است). حاصل لیستی از واژه‌های مشابه است.

نظرات خوانندگان

نویسنده: علی رادمان فر
تاریخ: ۱۶:۵ ۱۳۹۱/۰۹/۰۷

Spell Checker فارسی هم با این امکانات وجود داره؟!

نویسنده: وحید نصیری
تاریخ: ۱۶:۲۹ ۱۳۹۱/۰۹/۰۷

با فارسی هم کار می‌کنه: (، و یا)

نویسنده: بافکر
تاریخ: ۱۷:۷ ۱۳۹۳/۰۲/۳۱

با سلام

حالتی هست که اینجا هم بشه اینسرت، دیلیت و آپدیت کرد؟ (همانند ایندکس اصلی)

نویسنده: محسن تقی پور
تاریخ: ۱۴:۱۳ ۱۳۹۳/۰۶/۱۳

با سلام

ببخشید یه سوال داشتم میخوام بدونم Create SpellChecker Index رو چه مواقعی باید فراخوانی کنم ؟
موقعی که خبر درج شد و ادیت شد و پاک شد باید فراخوانی بشه ؟ یا تو یه بازه زمانی خواص مثلا هر 24 ساعت ایجاد بشه ؟
میخوام بدونم ایا میشه این ایندکس ها رو آپدیت کرد یا اینکه هر دفعه تو ایندکس اصلیم تغییری ایجاد شد باید از اول ساخته شه ؟

نویسنده: وحید نصیری
تاریخ: ۱۴:۲۲ ۱۳۹۳/۰۶/۱۳

این مورد خاص SpellChecker قابلیت به روز رسانی ندارد. هر بار باید از صفر انجام شود. روزی یکبار کافی است.