

اصطلاح نامه در شیرپوینت

(SharePoint Thesaurus Files)

تهیه کننده: رضا محمدی

(شرکت مونتکو ایران)



اصطلاح نامه چیست؟

فایل اصطلاح نامه در شیرپوینت ۲۰۰۷ قابلیت است که سرویس جستجوی شیرپینت را توسعه داده و کاربران را قادر می‌سازد تا هنگام جستجو برای یک عبارت نتایج را دریافت کنند که علاوه بر عبارت جستجو شده شامل واژه‌های مرتبط با آن عبارت نیز باشند. به عنوان مثال، جستجوی کلمه "خودرو" شامل نتایج باشد که حاوی کلمات "خودرو" و "ماشین" است. این حالت هنگامی میسر است که دو کلمه "ماشین" و "خودرو" در فایل اصطلاح نامه به یکدیگر مرتبط شده باشند. در این فایل، از مجموعه‌های جایگزین^۱ برای مشخص کردن واژه‌هایی که باید با یکدیگر جایگزین شوند و از مجموعه‌های گسترش^۲ برای مشخص کردن واژه‌های مترادف استفاده می‌شود.

فایل اصطلاح نامه

هنگامی که شیرپوینت ۲۰۰۷ نصب می‌شود، به ازای هر یک از زبان‌هایی که توسط شیرپوینت پشتیبانی می‌شود فایل اصطلاح نامه آن زبان تولید می‌شود. علاوه بر این فایل‌ها، فایل پیش فرضی به نام `tsneu.xml` نیز تولید می‌شود که مخصوص زبان‌هایی است که به ازای آن‌ها فایلی وجود ندارد. البته این فایل حالت عمومی داشته و همواره در جستجوها استفاده می‌شود حتی اگر عبارت جستجو به زبانی باشد که دارای فایل اصطلاح نامه باشد. برای اطلاعات بیشتر به بخش "فهرست فایل‌های اصطلاح نامه زبان‌ها" در انتهای این سند مراجعه کنید.

به طور پیش فرض، فایل‌های اصطلاح نامه بر روی سرور جستجو و در محل زیر نصب می‌شوند:

Drive:\ Program Files\Microsoft Office Servers\12.0\Data\Config

به ازای هر نمونه از سرویس جستجو بر روی سرور جستجو، فایل‌های اصطلاح نامه از این محل پیش فرض به پوشه زیر کپی می‌شوند:

Drive:\Program Files\Microsoft Office Servers\12.0\Data\Office
Server\Applications*<Application UID>*\Config

که در آن *<Application UID>* شناسه مربوط به (SSP) Shared Service Provider است.

¹ Replacement sets

² Expansion sets

نکته

در صورت ایجاد تغییر در فایل‌های اصطلاح نامه که در مسیر پیش فرض قرار دارند، با ایجاد هر SSP جدید، نسخه تغییر یافته آن‌ها به صورت خودکار کپی می‌شود اما اگر ایجاد تغییرات در فایل‌ها پس از ساختن SSP انجام شود، کپی کردن فایل‌های اصلاح شده از محل پیش فرض به پوشه مربوط به SSP باید دستی انجام پذیرد.

نکته مهم

در مسیری که فایل‌های اصطلاح نامه قرار دارد فایلی با نام `tsschema.xml` وجود دارد. این فایل را به هیچ عنوان تغییر ندهید. این فایل توسط همه فایل‌های اصطلاح نامه استفاده می‌شود و ایجاد تغییر در آن می‌تواند به اختلال در سرویس جستجو منجر شود.

به طور پیش فرض، محتوایی که به عنوان نمونه درون فایل‌های اصطلاح نامه قرار دارند، غیر فعال هستند. جهت قابل استفاده شدن این فایل‌ها برای سرویس جستجو، باید محتویات پیش فرض آن‌ها را ویرایش کرد.

فایل‌های اصطلاح نامه دارای دو نوع مدخل^۳ اصلی هستند:

- مجموعه‌های جایگزین^۴
- مجموعه‌های گسترش^۵

به جز دو مدخل اصلی فوق که در ادامه در مورد آن‌ها توضیح داده خواهد شد، مدخل دیگری وجود دارد به نام `diarictics_sensitive` که مقدار آن ۰ یا ۱ است. در صورتی که مقدار آن ۱ باشد سرویس جستجو، علائم مربوط به لهجه (اکسان) را در جستجو در نظر می‌گیرد (در زبان‌هایی مانند زبان فرانسه که دارای چنین خصوصیتی هستند) و در غیر این صورت آن علائم را نادیده می‌گیرد. مقدار پیش فرض این مدخل برابر ۰ است.

^۳ Entry

^۴ Replacement sets

^۵ Expansion sets



نمونه‌ای از یک فایل اصطلاح نامه به شکل زیر است:

```
<XML ID="Microsoft Search Thesaurus" >
<!-- Commented out
<thesaurus xmlns="x-schema:tsSchema.xml">
  <diacritics_sensitive>0</diacritics_sensitive>
  <expansion>
    <sub>Internet Explorer</sub>
    <sub>IE</sub>
    <sub>IE5</sub>
  </expansion>
  <replacement>
    <pat>NT5</pat>
    <pat>W2K</pat>
    <sub>Windows 2000</sub>
  </replacement>
  <expansion>
    <sub>run</sub>
    <sub>jog</sub>
  </expansion>
</thesaurus>
-->
</XML>
```

نکته مهم

تعداد آیتم‌هایی که در فایل اصطلاح نامه تعریف می‌شود در کارایی سیستم مؤثر است. توصیه می‌شود که تعداد آیتم‌ها بیش از ۱۰۰۰ مورد نشود. همچنین حداکثر تعداد آیتم‌های قابل تعریف ۱۰۰۰۰۰ مورد است. توجه داشته باشید که هر برچسب^۶ <pat> یا <sub> یک آیتم حساب می‌شود.

نکته

عباراتی که به فایل اصطلاح نامه اضافه می‌شوند نباید شامل کاراکترهای خاص باشند اما می‌توان از فضای خالی^۷ استفاده کرد. به عنوان مثال، اگر می‌خواهید جستجو برای واژه‌ای خاص، بدون نتیجه باشد از فضای خالی استفاده کنید. در مثال زیر جستجو برای عبارت "windows" نتیجه‌ای را بر نمی‌گرداند.

```
<replacement>
  <pat>windows</pat>
  <sub></sub>
</replacement>
```

Tag^۶
Blank^۷



استفاده از مجموعه‌های جایگزین

یک مجموعه جایگزین معرف مجموعه‌ای است که در آن یک الگو در یک عمل جستجو توسط یک یا چند عبارت دیگر، جایگزین می‌شود. به عنوان مثال، می‌توان مجموعه‌ای داشت که در آن عبارت "W2K" با عبارت "Windows 2000" جایگزین شود. در این حالت هنگامی که کاربر واژه "W2K" را جستجو می‌کند، تنها نتایجی به وی نمایش داده خواهد شد که شامل عبارت "Windows 2000" باشند و این نتایج شامل عبارت "W2K" نخواهند بود.

هر مجموعه جایگزین در درون برچسبی به نام <replacement> قرار می‌گیرد. در درون این برچسب، تعدادی الگو در درون برچسب‌های <pat> قرار می‌گیرند و یک یا چند عبارت جایگزین در درون برچسب <sub> قرار می‌گیرند. الگوها و عبارات جایگزین می‌توانند شامل یک کلمه یا مجموعه‌ای از کلمات باشند. به عنوان مثال برای افزودن یک مجموعه جایگزین که در آن باید الگوی "W2K" با عبارت "Windows 2000" جایگزین شود به صورت زیر عمل می‌کنیم:

```
< replacement >  
  < pat >W2K < / pat >  
  < sub >Windows 2000 < / sub >  
< / replacement >
```

توجه داشته باشید که برای هر الگو می‌توان بیش از یک جایگزین داشت.

استفاده از مجموعه‌های گسترش

یک مجموعه گسترش گروهی است متشکل از عبارات مترادف. هنگامی که یکی از عبارات این گروه مورد پرس‌وجو قرار می‌گیرد، عبارات مترادف با آن نیز در جستجو لحاظ شده و در نتایج نشان داده می‌شوند. به عنوان مثال، اگر مجموعه گسترشی شامل عبارات هم معنی زیر داشته باشیم:

- Writer
- author
- journalist

اگر واژه "author" مورد جستجو قرار گیرد، موتور جستجوی شیرپوینت نتایجی که شامل دو عبارت "writer" و "journalist" هم باشد برمی‌گرداند.

هر مجموعه گسترش در درون برچسبی به نام <expansion> قرار می‌گیرد. در درون برچسب <expansion> یک یا چند عبارت در درون برچسب <sub> قرار می‌گیرد. به عنوان مثال، برای نمونه بالا خطوط زیر به فایل اصطلاح نامه اضافه می‌شود:

```
<expansion>
  <sub>writer</sub>
  <sub>author</sub>
  <sub>journalist</sub>
</expansion>
```

در یک فایل اصطلاح نامه می‌توان واژه‌های منفرد یا عبارات ترکیبی داشت. واژه‌شکنی^۸ که برای یک زبان وجود دارد بر اساس قوانین لغوی آن زبان محدوده شکستن عبارات را تشخیص می‌دهد. اگر واژه‌ای در فایل اصطلاح نامه وجود داشته باشد که واژه‌شکن نتواند آن را به عنوان یک واژه منفرد تشخیص دهد باید آن را در یک لغت‌نامه سفارشی^۹ نیز اضافه کرد تا واژه‌شکن آن واژه را به بخش‌های کوچکتر تقسیم نکند. به عنوان مثال، اگر در یک مجموعه گسترش، واژه "IT&T" وجود داشته باشد اما در لغت‌نامه سفارشی اضافه نشده باشد، واژه‌شکن ممکن است آن را به دو بخش "IT" و "T" بشکند و در نتیجه حاصل جستجو برای واژه "IT&T" به نتایج مورد انتظار منجر نشود. برای اطلاعات بیشتر در مورد ایجاد و نحوه استفاده از یک لغت‌نامه سفارشی به آدرس زیر مراجعه کنید:

<http://technet.microsoft.com/en-us/library/cc263242.aspx>

ویرایش یک فایل اصطلاح نامه

برای ویرایش یک فایل اصطلاح نامه به صورت زیر عمل کنید:

۱. فایل اصطلاح نامه را در برنامه Notepad باز کنید.
۲. اگر فایل مورد نظر برای نخستین بار ویرایش می‌شود ابتدا برچسب‌های <!--!> و --> را که در ابتدا و انتهای فایل قرار دارند را حذف کنید تا بخش‌های غیرفعال شده فعال شوند.
۳. تغییرات مورد نظر را با افزودن، اصلاح و یا حذف مجموعه‌های جایگزین و گسترش اعمال کنید.
۴. فایل اصطلاح نامه را ذخیره کنید.
۵. سرویس Office SharePoint Server Search را راه اندازی مجدد کنید.



فهرست فایل‌های اصطلاح نامه زبان‌ها

زبان	نام فایل
Arabic	tsara.xml
Bengali	tsben.xml
Bulgarian	tsbul.xml
Catalan	tscat.xml
Chinese (Simplified)	tschs.xml
Chinese (Traditional)	tscht.xml
Croatian	tscro.xml
Dutch (Netherlands)	tsnld.xml
English (United Kingdom)	tseng.xml
English (United States)	tsenu.xml
Finnish	tsfin.xml
French	tsfra.xml
German	tsdeu.xml
Gujarati	tsguj.xml
Hebrew	tsheb.xml
Hindi	tshin.xml
Icelandic	tsice.xml
Indonesian	tsind.xml
Italian	tsita.xml
Japanese	tsjpn.xml
Kannada	tskan.xml
Korean	tskor.xml
Latvian	tslat.xml
Lithuanian	tslit.xml
Malay	tsmal.xml
Malayalam	tsmly.xml
Marathi	tsma.xml
Language Neutral	tsneu.xml
Norwegian (Bokmal)	tsnor.xml
Polish	tsplk.xml
Polish	tspol.xml
Portuguese (Brazil)	tsptb.xml
Portuguese (Portugal)	tspor.xml
Punjabi	tspun.xml
Romanian	tsrom.xml
Russian	tsrus.xml
Serbian (Cyrillic)	tssbc.xml
Serbian (Latin)	tssbl.xml
Slovak	tssvk.xml
Slovenian	tsslo.xml
Spanish	tsesn.xml
Swedish	tssve.xml
Tamil	tstam.xml
Telugu	tstel.xml
Thai	tstha.xml
Turkish	tstur.xml
Ukrainian	tsukr.xml
Urdu (Pakistan)	tsurd.xml