

عبارت باقاعده (Regular Expressions) در دات نت

امیدوارم این مقاله برای دوستان مفید باشد ...

عبارت با قاعده Regular expressions تا کنون در ابزارها و زبان های برنامه نویسی متعددی به یکارگرفته شده است. از جمله موارد آشنا در مورد کاربرد عبارت با قاعده می توان به اعتبار سنجی آدرس پست الکترونیک ، کلمه رمز ، .. اشاره نمود. در سکوی دات نت هم که یکی از جدیدترین فناوریهای توسعه نرم افزاری می باشد فضای نام و مجموعه ای از کلاسهای متعدد برای بهره گیری از قدرت عبارت با قاعده موجود می باشد .

کلاسهای عبارت باقاعده موجود در چارچوب دات نت قسمتی از کتابخانه کلاس پایه می باشند که در این صورت هر زبان یا ابزاری که به زبان مشترک زمان اجرا (CLI) رجوع کند می تواند از این کلاس ها استفاده نماید که در این مقاله نمونه مثال هایی با زبان شی گرا و قدرتمند (C#.Net) ارائه می شود..

1. مقدمه :

لازم است ابتدا برای درک درست مبحث در مورد بعضی از موارد پایه ای توضیح داده شود با توجه به یک تعریف رسمی عبارت با قاعده عبارتی است که بتوان آن را از روی حروف الفبا و رشته تهی و با استفاده از عملگرهای اتحاد ، اتصال ، * یا + ایجاد کرد همانند

you (@) somewhere.com و یا aabbaa

در دات نت عبارت باقاعده عبارتی است که از کاراکتر های یونی کد، با استفاده از عملگرهای ؟ ، + ، * ، | و .. ایجاد شود .

عبارت با قاعده روش های قدرتمند ، راحت و کارا برای پردازش متن ارائه می کنند. عبارت با قاعده به ما اجازه می دهد برای پیدا کردن الگوی کاراکتری خاص به سرعت مقدار زیادی از متن را تجزیه نموده و یا رشته ای را از متنی استخراج، ویرایش ، جایگذاری یا حذف نموده و یا حتی امکان اضافه کردن رشته استخراج شده از متن به یک مجموعه برای تولید گزارش را به ما میدهد. در خیلی از برنامه هایی که با رشته ها ارتباط دارند

همچون پردازش فایل HTML و یا تجزیه فایل log یا تجزیه عنوان (HTTP عبارات با قاعده ابزاری ضروری) می باشند.

عبارت با قاعده درون چارچوب دات نت بیشترین ویژگی های مورد پسند در دیگر زبانها همچون پرل (Perl) و.. را دارا بوده ضمن اینکه شامل دیگر خصوصیات تقریباً منحصر بفرد همچون تطبیق راست به چپ و کامپایل در هنگام اجرا نیز می باشد.

می توان عبارت با قاعده را به عنوان زبانی فرض نمود که برای دستکاری در متون طراحی و بهینه سازی شده است این زبان شامل دو نوع کاراکتر ابتدایی می باشد:

کاراکتر های حرفی (Literal) و فوق کاراکتر ها که قدرت پردازش را به عبارت باقاعده میدهند.

احتمالاً شما با فوق کاراکتر های ؟ و * که در سیستم فایلی داس برای نمایش یک کاراکتر یا گروهی از کاراکتر ها مورد استفاده قرار می گرفت(می گیرد) آشنایی دارید به طور مثال در دستور زیر که وظیفه آن کپی هر گونه فایل موجود در درایو C: با پسوند .doc بر روی دیسکت موجود در درایو A: می باشد ، فوق کاراکتر * به عنوان هر گونه فایلی در نظر گرفته می شود که دارای پسوند .doc می باشد.

عبارت با قاعده این ایده ابتدایی را چندین برابر توسعه داده اند و یک دسته بزرگ از فوق کاراکترها را ارائه کرده اند که امکان توصیف عبارت پیچیده برای تطابق متون با کاراکترهای نسبتاً کمتر فراهم کرده اند.

به طور مثال : وقتی عبارت باقاعده برای بدنه یک متن به کار می رود تمامی رشته های "s2000\ " که با هر کاراکتر فاصله مثل space یا tab موجود می باشد را تطبیق داده و مشخص میکند. البته برای زبان سی شارپ (و ++c, jscrip) می بایست این عبارت را به شکل "s2000\\" بنویسیم.

عبارات با قاعده اغلب اوقات براي اين مورد استفاده قرار مي گيرند که مشخص کنند آیا یک رشته با یک الگوي مشخص تطابق دارد یا نه ، به طور مثال آیا رشته مورد نظر شبیه به کد پستي یا شماره تلفن یا آدرس ایمیل مشخصي مي باشد یا نه. عبارتي اعتبار سنجي يکي از کاربردهاي اساسي عبارات با قاعده مي باشد.

يکي از ساده ترين عبارات با قاعده، زیر رشته مي باشد براي نمونه عبارت باقاعده "فارس" با رشته "خليج فارس" تطابق دارد چونکه اين رشته شامل عبارت با قاعده مي باشد.

مي توان به راحتی رشته هايي که با کاراکتر هاي خاصي شروع يا پايان مي يابند را مشخص نمود: کاراکتر ^ شروع رشته را تطبیق میدهد به طور مثال good^ با رشته هاي good morning , good evening , ... تطبیق میکند.

براي تطبیق انتهاي رشته از کاراکتر \$ استفاده میکنیم به طور مثال \$ere با رشته هاي Where , There مطابقت میکند.

دسته کاراکترها [Character Classes] به ما اجازه میدهد مجموعه اي از کاراکترها را مشخص کنیم به طور مثال دسته [aeiou] با رشته "Hey" تطابق دارد ولي با "ZZz" نه به عبارت دیگر رشته باید حداقل شامل يکي از کاراکتر هاي موجود در دسته باشد.

با استفاده از دسته کاراکترها همچنين ميتوانيم یک محدوده اي را مشخص کنیم به عنوان نمونه [۰-۹] با اعداد ۰, ۵, ۸ و به طور کلي هر عدد در اين محدوده ۰ تا ۹ تطابق خواهد داشت.

بعضي مواقع مي خواهيم یک تعداد عدد مشخص را با یک الگوي مشخصي تطبیق دهيم براي مثال یک کد پستي ۵ رقمي را در نظر بگيريد بنابراين به اين شکل نوشته مي شود :

$\$ \{5\} \backslash d \{5\} \wedge$ يا $\{5\} [0-9] \wedge$

توجه : هنگام تايپ و خواندن از راست به چپ مي باشد

يعني ابتدای رشته را تطبیق دهد ، با ۵ رقم دنبال کند و بالاخره انتهاي رشته را دنبال نماید به طور مثال اين عبارت باقاعده با عدد ۱۲۳۴۵ که به طور فرضي یک کد پستي است تطابق دارد. و يا عبارت $\wedge [0-9] \{5\} \$$ با یک کد ۱۰ رقمي مثل ۱۲۳۴۵-۰۹۸۷۶ تطابق خواهد داشت. دستور زبان یک عبارت باقاعده شامل تعدادي خصیصه مي باشد که مي توان با هم ترکیب نمود به طور مثال ميتوان بوسيله (?! :) خاصیت غير حساس بودن به بزرگ و کوچک بودن حروف را فعال نمود در نتیجه عبارت (?!:[aeiou]) با رشته هاي hello و HELLO تطابق خواهد داشت ولي با ZZZZ نه.

در صورت علاقه دوستان ادامه خواهد داشت....

- 2- درک دستور زبان ابتدایي عبارات با قاعده در دات نت
- 3- معرفي کلاس هاي عبارات با قاعده
- 4- تعيين اعتبار ورودی کاربر بوسيله عبارات با قاعده
- 5- تعویض زیر رشته ها با استفاده از عبارات با قاعده
- 6- معرفي کنترل RegularExpressionValidator

نویسنده Hamid.R Parvareh

کليه حقوق این مقاله متعلق به نویسنده و سایت برنامه نویسی مي باشد .
استفاده از مطالب این مقاله در صورت ذکر منبع مجاز است .

www.barnamenevis.org