عنوان: **حذف نمودن کاراکتر های ناخواسته توسط Recursive CTE قسمت اول**
نویسنده: محمد سلیم ابادی
تاریخ: ۱۳۹۱/۱۱/۰۱ ۱۵:۳۵
آدرس: www.dotnettips.info
برچسب‌ها: SQL Server, T-SQL, recursive cte, replace

شاید برایتان تا حالا پیش آمده باشد که بخواهید یکسری کاراکترهای ناخواسته و اضافه را از یک رشته حذف کنید. بطور مثال تمام کاراکتر هایی غیر عددی را باید از یک رشته حذف نمود تا آن رشته قابلیت تبدیل به نوع integer را بدست بیاورد.

اگر تعداد کاراکترهای ناخواسته محدود و مشخص هستند می‌توانید با دستور REPLACE آنها را حذف کنید، مثلا میخواهیم هر سه کاراکتر ~!@ از رشته حذف شوند:

```
DECLARE @s VARCHAR(50) = '~~~~~~!@@@@@@ salam';
SET @s = REPLACE(REPLACE(REPLACE(@s, '~', ''), '!', ''), '@', '');
SELECT @s AS new_string
```

ولی هنگامی که کاراکترها نامحدود بوده امکان نوشتن تابع REPLACE به کرات بی معنا است در این حالت باید دنبال روشی پویا و تعمیم پذیر بود.
با جستجویی که در اینترنت انجام دادم متوجه شدم تکنیک WHILE یا همون loop یکی از روش‌های رایج برای انجام اینکار هست، که احتمالا به دلیل سهولت در بکارگیری و سادگی آن بوده که عمومیت پیدا کرده است.
مستقل از این صحبتها هدف معرفی یک روش مجموعه گرا (set-based) برای این مساله می‌باشد.

حذف کاراکترها ناخواسته با تکنیک Recursive CTE
راه حل بر اساس جدول زیر است:

```
CREATE TABLE test_string
(id integer not null primary key,
 string_value varchar(500) not null);

INSERT INTO test_string
VALUES (1, '@@@@ #### salam 12345'),
       (2, 'good $$$$$ &&&&&& bye 00000');
```

حالا فرض کنید می‌خواهیم هر کاراکتری غیر از حروف الفبای انگلیسی و فاصله(space) از رشته حذف شود. پس دو داده فوق به صورت salam و good bye در انتها در خواهند آمد.
برای حذف کاراکترهای ناخواسته فوق query زیر را اجرا کنید.

```
WITH CTE (ID, MyString, Ix) AS
(
    SELECT id,
           string_value,
           PATINDEX('%[^a-z ]%', string_value)
    FROM   test_string

    UNION ALL

    SELECT id,
           CAST(REPLACE(MyString, SUBSTRING(MyString,Ix , 1), '') AS VARCHAR(500)),
           PATINDEX('%[^a-z ]%', REPLACE(MyString, SUBSTRING(MyString,Ix , 1), ''))
    FROM   CTE
    WHERE  Ix > 0
)
SELECT *
FROM   cte
--WHERE  Ix = 0;
ORDER BY id, CASE WHEN Ix = 0 THEN 1 ELSE 0 END, Ix;
```

توضیح query:

در قسمت anchor اندیس اولین کاراکتر ناخواسته (خارج از رنج حروف الفبا و فاصله) بدست می‌آید. سپس در قسمت recursive هر کاراکتری که برابر باشد با کاراکتر ناخواسته ای که در مرحله قبل بدست آمده از رشته حذف می‌شود این عملیات توسط تابع replace صورت می‌گیرد و اندیس کاراکتر ناخواسته بعدی بعد از حذف کاراکتر ناخواسته قبلی بدست می‌آید که به مرحله بعد منتقل می‌شود. این مراحل تا آنجایی پیش می‌رود که دیگر کاراکتر ناخواسته ای در رشته وجود نداشته باشد.

به جدول زیر توجه بفرمایید (خروجی query فوق)

	ID	MyString	lx
1	1	@@@@ #### salam 12345	1
2	1	#### salam 12345	2
3	1	salam 12345	9
4	1	salam 2345	9
5	1	salam 345	9
6	1	salam 45	9
7	1	salam 5	9
8	1	salam	0
9	2	good $$$$$ &&&&&& bye 00000	6
10	2	good &&&&&& bye 00000	7
11	2	good bye 00000	12
12	2	good bye	0

نتیجه مطلوب ما آن دو سطری است که در کادر بنفش هستند. که اگر به ستون Ix اشان توجه کنید مقدارش برابر با 0 است.

لطفا به سطر اول جدول توجه بفرمایید مشاهده می‌شود که هر 4 کاراکتر @ یکباره از رشته حذف شدند که بدلیل استفاده از تابع REPLACE میباشد.

نظرات خوانندگان

نویسنده: سید حمزه
تاریخ: ۱۳۹۲/۰۶/۲۷ ۱۶:۵۳

سلام
خیلی جامع بود
فقط من متوجه نشدم دقیقا کجای کوئریم باید اینو بنویسم.
و همینطور میخواستم اگر بشه یک فانکشن بسازم و از اون هم توی select استفاده کنم.
ممنون