

How do I count the number of times a word or phrase occurs in a string?

Assuming that you don't have a built-in function that does this in one shot...

Here is a trick for quickly (i.e. using minimal typing effort) counting the number of times a string occurs in a larger string:

Below is some pseudocode for searching for a string called "needle" in a larger string called "haystack" :

```
numOccur = (length(haystack) - length(replace(haystack, needle, ""))) / length(needle)
```

note:

replace(string a, string b, string c) modifies a by replacing each occurrence of b with c

Why does this work? What we're doing is finding the difference in length of the haystack after we remove all occurrences of the needle. If we divide this difference by the length of the needle, then we get the number of needles that were removed. This is equivalent to the number of occurrences of the needle in the haystack.

An Example:

The following is a SQL query for PostgreSQL that puts this to use. This analyzes a collection of article titles and abstracts and counts the total number of occurrences of each phrase in a list of phrases:

```
SELECT a.phrase, sum((length(b.title) - length(replace(b.title, a.phrase, ''))) / length(a.phrase)) as numOccur
```

Revision #10

Created Wed, Aug 9, 2006 1:33 AM by Fong, Jason

Updated Thu, Sep 21, 2006 12:51 PM by Fong, Jason