strcmp of String Concatenation and String Repetition

Yi Tang

April 5, 2019
(Last updated on January 31, 2022)

In this note we prove a property regarding strcmp, string concatenation, and string repetition. The problem is put forward in some private chat, and is said to originate as an interview question.

We start with introducing the basic definitions of strings and string operations.

Definition 1. A string a over alphabet Σ is a finite-length sequence over Σ, i.e., $a \in \Sigma^* := \bigcup_{n=0}^{\infty} \Sigma^n$. We fix some alphabet Σ (that is equipped with a total order) throughout this note and often omit it for simplicity. For strings a, b over the same alphabet, we have the following operations.

- The string concatenation ab is defined to be the concatenated sequence.
- For a number $k \in \mathbb{N}$, the string repetition a^k is recursively defined to be $a^{k-1}a$ if $k > 0$ and the empty string ε if $k = 0$.
- The string length $|a|$ is defined to be the length of a as a sequence.
- For indices $i \le j \le |a| \in \mathbb{N}$, the substring $a_{i:j}$ is defined to be the corresponding subsequence (including i, excluding j, and counting from 0); we also define $a_{:j} := a_{0:j}$ and $a_{i:} := a_{i:|a|}$.
- If the alphabet is totally ordered, then there exists a natural *lexicographic order* among the strings. The lexicographic order indicator $\texttt{strcmp}(a, b)$ is defined to be -1 if $a < b$, 0 if $a = b$, and $+1$ if $a > b$. Note that $\texttt{strcmp}(a, b) = 1[a > b] - 1[a < b] = 1[a \ge b] - 1[a \le b]$.[1]

We now state the main theorem of this note regarding strcmp, string concatenation, and string repetition.

Theorem 2 (strcmp of string concatenation and string repetition)**.** *For any strings a, b,*

$$\texttt{strcmp}(ab, ba) = \texttt{strcmp}(a^{|b|}, b^{|a|}) \ .$$

To prove Theorem 2, we first show some handy properties of strcmp. The proofs for these properties are straightforward by the lexicographic order and are thus omitted.

Lemma 3 ("Transitivity" of strcmp)**.** *For any strings a, b, c, if $\texttt{strcmp}(a, b) = \texttt{strcmp}(b, c)$, then*

$$\texttt{strcmp}(a, b) = \texttt{strcmp}(b, c) = \texttt{strcmp}(a, c) \ .$$

Lemma 4 (Decomposition of strcmp)**.** *For any strings a, b and index $k \le \min(|a|, |b|)$,*

$$\texttt{strcmp}(a, b) = \texttt{strcmp}(a_{:k}, b_{:k}) \parallel \texttt{strcmp}(a_{k:}, b_{k:}) \ ,$$

where $x \parallel y$ denotes the generalized logical operation "if $x \ne 0$ then x otherwise y".

[1] This suggests that we can state properties of strcmp in terms of $\ge$ which is less fancy, while nevertheless we try to present all statements as well as proofs using strcmp, showing that it is possible to deal with strcmp formally and concisely.

Corollary 5 (Padding of strcmp). *For any strings a, b, c,*

$$\texttt{strcmp}(a, b) = \texttt{strcmp}(ca, cb) \ .$$

Furthermore, if $|a| = |b|$, then

$$\texttt{strcmp}(a, b) = \texttt{strcmp}(ac, bc) \ .$$

The corollary immediately follows from Lemma 4, while we include a proof here to demonstrate how calculations can be done formally using the generalized logical operator $\|$.

Proof. By Lemma 4,

$$\texttt{strcmp}(ca, cb) = \texttt{strcmp}(c, c) \parallel \texttt{strcmp}(a, b) = 0 \parallel \texttt{strcmp}(a, b) = \texttt{strcmp}(a, b) \ ,$$

and when $|a| = |b|$,

$$\texttt{strcmp}(ac, bc) = \texttt{strcmp}(a, b) \parallel \texttt{strcmp}(c, c) = \texttt{strcmp}(a, b) \parallel 0 = \texttt{strcmp}(a, b) \ .$$

□

We also show the following property of strcmp involving string functions.

Lemma 6 ("Conditional" strcmp). *For any strings a, b and string functions f, g,*

$$\texttt{strcmp}(a, b) \parallel \texttt{strcmp}(f(a), g(a)) = \texttt{strcmp}(a, b) \parallel \texttt{strcmp}(f(b), g(b)) \ .$$

Proof. If $\texttt{strcmp}(a, b) \neq 0$, then both the left- and right-hand sides are equal to $\texttt{strcmp}(a, b)$. Otherwise, $\texttt{strcmp}(a, b) = 0$ implies $a = b$, and thus the left- and right-hand sides are equal as well. □

Now we present the proof of Theorem 2, the main theorem of this note.

Proof of Theorem 2. Our goal is to prove

$$\texttt{strcmp}(ab, ba) = \texttt{strcmp}(a^{|b|}, b^{|a|}) \ . \tag{1}$$

Without loss of generality suppose $|a| \leq |b|$, and let $b_1 := b_{:|a|}$ and $b_2 := b_{|a|:}$. We prove by induction on $|a| + |b|$. Note that Equation (1) obviously holds when $a = \varepsilon$ or $b = \varepsilon$, i.e., $|a| + |b| \leq 1$. Now assume the inductive hypothesis that Equation (1) holds for all a, b such that $|a| + |b| \leq k$ $(k \geq 1)$. Then for $|a| + |b| = k + 1$, by Lemmas 4 and 6,

$$\begin{aligned}
\texttt{strcmp}(ab, ba) &= \texttt{strcmp}(a, b_1) \parallel \texttt{strcmp}(b_1 b_2, b_2 a) \\
&= \texttt{strcmp}(a, b_1) \parallel \texttt{strcmp}(a b_2, b_2 a) \ ; \\
\texttt{strcmp}(a^{|b|}, b^{|a|}) &= \texttt{strcmp}(a^{|b_1|+|b_2|}, (b_1 b_2)^{|a|}) \\
&= \texttt{strcmp}(a, b_1) \parallel \texttt{strcmp}(a^{|b_1|+|b_2|-1}, (b_2 b_1)^{|a|-1} b_2) \\
&= \texttt{strcmp}(a, b_1) \parallel \texttt{strcmp}(a^{|b_2|+|a|-1}, (b_2 a)^{|a|-1} b_2) \ .
\end{aligned}$$

Hence it suffices to prove

$$\texttt{strcmp}(a b_2, b_2 a) = \texttt{strcmp}(a^{|b_2|+|a|-1}, (b_2 a)^{|a|-1} b_2) \ . \tag{2}$$

If $|a| = 1$, then $\texttt{strcmp}(a^{|b_2|+|a|-1}, (b_2 a)^{|a|-1} b_2) = \texttt{strcmp}(a^{|b_2|}, b_2^{|a|})$, which implies Equation (2) by inductive hypothesis. When $|a| > 1$, note that for all $0 \leq j \leq i < |a| - 1$, by Corollary 5,

$$\begin{aligned}
\texttt{strcmp}(a b_2, b_2 a) = \texttt{strcmp}(&a^i b_2^{i-j} (a b_2) b_2^j (b_2 a)^{|a|-i-2} b_2, \\
&a^i b_2^{i-j} (b_2 a) b_2^j (b_2 a)^{|a|-i-2} b_2) \ ,
\end{aligned}$$

which gives a transition from $a^{|a|-1}b_2^{|a|}$ $(i = j = |a| - 2)$ to $(b_2a)^{|a|-1}b_2$ $(i = j = 0)$. Therefore by Lemma 3,

$$\texttt{strcmp}(ab_2, b_2a) = \texttt{strcmp}(a^{|a|-1}b_2^{|a|}, (b_2a)^{|a|-1}b_2) \ . \tag{3}$$

Moreover, by inductive assumption and Corollary 5,

$$\begin{aligned}
\texttt{strcmp}(ab_2, b_2a) &= \texttt{strcmp}(a^{|b_2|}, b_2^{|a|}) \\
&= \texttt{strcmp}(a^{|a|-1}a^{|b_2|}, a^{|a|-1}b_2^{|a|}) \\
&= \texttt{strcmp}(a^{|b_2|+|a|-1}, a^{|a|-1}b_2^{|a|}) \ .
\end{aligned}$$

Combining with Equation (3) and using Lemma 3, we get Equation (2), which concludes the proof. □