Answers to the mid-term exam on Compilers

Christian Rinderknecht

25 October 2005

Question 1. Let the alphabet $\Sigma = \{a, b\}$ and the following regular expressions:

$$r = a(a\,|\,b)^{\star}ba$$
$$s = (ab)^{\star}\,|\,(ba)^{\star}\,|\,(a^{\star}\,|\,b^{\star})$$

The language denoted by r is noted $L(r)$ and the language denoted by s is noted $L(s)$.

Find a word x such as

1. $x \in L(r)$ and $x \notin L(s)$
2. $x \notin L(r)$ and $x \in L(s)$
3. $x \in L(r)$ and $x \in L(s)$
4. $x \notin L(r)$ and $x \notin L(s)$

Answer 1. The method to answer these questions is simply to try small words by constructing them in order to satisfy the constraints.

1. The shortest word x belonging to L(r) is found by taking ε in place of $(a|b)^{\star}$. So $x = aba$. Let us check if $x \in L(s)$ or not. $L(s)$ is made of the union of four sub-languages (subsets). To make this clear, let us remove the useless parentheses on the right side:

   $$s = (ab)^{\star}\,|\,(ba)^{\star}\,|\,a^{\star}\,|\,b^{\star}$$

   Therefore, membership tests on $L(s)$ have to be split into four: one membership test on $(ab)^{\star}$, one on $(ba)^{\star}$, one on $a^{\star}$ and another one on $b^{\star}$. In other words:

   $$x \in L(s) \Leftrightarrow x \in L((ab)^{\star}) \textit{ or } x \in L((ba)^{\star}) \textit{ or } x \in L(a^{\star}) \textit{ or } x \in L(b^{\star})$$

   Let us test the membership with $x = aba$:

   (a) The words in $L((ab)^{\star})$ are ε, ab, $abab\ldots$ Thus $aba \notin L((ab)^{\star})$.

(b) The words in $L((ba)^\star)$ are ε, ba, $baba\ldots$ Hence $aba \notin L((ba)^\star)$.

(c) The words in $L(a^\star)$ are ε, a, $aa\ldots$ Therefore $aba \notin L(a^\star)$.

(d) The words in $L(b^\star)$ are ε, b, $bb\ldots$ So $aba \notin L(b^\star)$.

Finally the conclusion is $aba \notin L(s)$, which is what we were looking for.

2. What is the shortest word belonging to $L(s)$? Since the four sub-languages composing $L(s)$ are starred, it means that $\varepsilon \in L(s)$. Since we showed at the item (1) that aba is the shortest word of $L(r)$, it means that $\varepsilon \notin L(r)$ because ε is of length 0.

3. This question is a bit more difficult. After a few tries, we cannot find any x such as $x \in L(r)$ and $x \in L(s)$. Then we may try to prove that $L(r) \cap L(s) = \varnothing$, i.e. there is no such x. How should we proceed? The idea is to use the decomposition of $L(s)$ into for sub-languages and try to prove

$$\begin{aligned} L(r) \cap L((ab)^\star) &= \varnothing \\ L(r) \cap L((ba)^\star) &= \varnothing \\ L(r) \cap L(a^\star) &= \varnothing \\ L(r) \cap L(b^\star) &= \varnothing \end{aligned}$$

Indeed, if all these four equations are true, they imply $L(r) \cap L(s) = \varnothing$.

(a) Any word in $L(r)$ finishes with a whereas any word in $L((ab)^\star)$ finishes with b or is ε. Thus $L(r) \cap L((ab)^\star) = \varnothing$.

(b) For the same reason, $L(r) \cap L(b^\star) = \varnothing$.

(c) Any word in $L(r)$ contains both a and b whereas any word in $L(a^\star)$ contains only b or is ε. Therefore $L(r) \cap L(a^\star) = \varnothing$.

(d) Any word in $L(r)$ starts with a whereas any word in $L((ba)^\star)$ starts with b or is ε. Thus $L(r) \cap L((ba)^\star) = \varnothing$.

Finally, since all the four equations are false, they imply that

$$L(r) \cap L(s) = \varnothing.$$

4. Let us construct letter by letter a word x which does not belong neither to $L(r)$ not $L(s)$. First, we note that all words in $L(r)$ start with a, so we can try to start x with b: this way $x \notin L(r)$. So we have $x = b\ldots$ and we have to fill the dots with some letters in such a way that $x \notin L(s)$.

We use again the decomposition of $L(s)$ into four sub-languages and make sure that x does not belong to any of those sub-languages.

First, because x starts with a, $x \notin L(b^\star)$ and $x \notin L((ba)^\star)$.

Now, we have to add some more letters such as $x \notin L(a^\star)$ and $x \notin L((ab)^\star)$.

Since any word in $L(a^\star)$ has a a as second letter or is ε, we can choose the second letter of x to be b. This way $x = ab\ldots \notin L(a^\star)$.

Finally, we have to add more letters to make sure that

$$x = ab\ldots \notin L((ab)^\star)$$

Any word in $L((ab)^\star)$ is either ε or ab or $abab\ldots$, hence the third letter is a. Therefore, let us choose b as the third letter of x and we thus have $x = aba \notin L((ab)^\star)$. Summary:

$$aba \notin L(r)\ aba \notin L(b^\star)\ aba \notin L((ba)^\star)\ aba \notin L(a^\star)\ aba \notin L((ab)^\star)$$

which is equivalent to

$$aba \notin L(r) \text{ and } aba \notin L((ab)^\star) \cup L((ba)^\star) \cup L(a^\star) \cup L(b^\star) = L(s)$$

Therefore $x = aba$ is one possible answer.

Question 2. Given the binary alphabet $\Sigma = \{a, b, c\}$ and the order on letters $a < b < c$, write regular definitions for the following languages.

1. All words starting and ending with c.
2. All words in which the third last letter is b.
3. All words containing exactly three a.
4. All words containing at least one b before a c.
5. All words in which the letters are in increasing order.

Answer 2.

1. The constraint on the words is that they must be of the shape $c\ldots c$ where the dots stand for "any combination of a, b and c." In other words, one answer is $c\,(a\,|\,b\,|\,c)^\star\,c\,|\,c$.

2. The question implies that the words we are looking for are of the form $\ldots b\_\_$ where the dots stand for "any sequence of a, b and c" and each $\_$ stands for a regular expression denoting any letter. Any letter is described by $(a\,|\,b\,|\,c)$; therefore one possible answer is

$$(a\,|\,b\,|\,c)^\star b\,(a\,|\,b\,|\,c)\,(a\,|\,b\,|\,c)$$

3. The words we search contain, at any place, exactly three a, so are of the form $\ldots a \ldots a \ldots a \ldots$, where the dots stand for "any letter except a", i.e., "any number of b or c." In other words:

$$(b\,|\,c)^\star a\,(b\,|\,c)^\star a\,(b\,|\,c)^\star a\,(b\,|\,c)^\star$$

4. The words we search are of the form $\dots b \dots c \dots$, where the dots stand for "All words possible words made of a, b or c." Therefore it is easy to understand that a short answer is

$$(a\,|\,b|c)^\star b\,(a\,|\,b|c)^\star c\,(a\,|\,b|c)^\star$$

5. Because the alphabet is made only of two letters, the answer is easy: we put first all the a, then all the b and finally all the c:

$$a^\star b^\star c^\star$$

Question 3. Simplify, if possible, the following regular expressions.

$$(\varepsilon\,|\,a^\star\,|\,b^\star\,|\,a\,|\,b)^\star$$
$$a(a\,|\,b)^\star b\,|\,(ab)^\star\,|\,(ba)^\star$$

Answer 3.

1. The first regular expression can be simplified in the following way:

$$\begin{aligned}
(\varepsilon\,|\,a^\star\,|\,b^\star\,|\,a\,|\,b)^\star &= (\varepsilon\,|\,a^\star\,|\,b^\star\,|\,b)^\star && \textit{since}\, L(a) \subset L(a^\star)\\
&= (\varepsilon\,|\,a^\star\,|\,b^\star)^\star && \textit{since}\, L(b) \subset L(b^\star)\\
&= (\varepsilon\,|\,a^+\,|\,b^+)^\star && \textit{since}\, \{\varepsilon\} \subset L(x^\star)\\
&= (a^+\,|\,b^+)^\star && \textit{since}\, (\varepsilon\,|\,x)^\star = x^\star
\end{aligned}$$

Words in $L((a^+\,|\,b^+)^\star)$ are of the form $(a\dots a)(b\dots b)(a\dots a)(b\dots b)\dots$, so we recognise $(a\,|\,b)^\star$. Therefore $(\varepsilon\,|\,a^\star\,|\,b^\star\,|\,a\,|\,b)^\star = (a\,|\,b)^\star$.

2. The second regular expression can be simplified in the following way. We note first that the expression is made of the disjunction of three regular sub-expressions (i.e. it is a union of three sub-languages). The simplest idea is then to check whether one of these sub-languages is redundant, i.e. if one is included in another. If so, we can simply remove it from the expression.

$$\begin{aligned}
a(a\,|\,b)^\star b\,|\,(ab)^\star\,|\,(ba)^\star &= a(a\,|\,b)^\star b\,|\,\varepsilon\,|\,(ab)^+\,|\,(ba)^\star && \textit{since}\, (ab)^\star = \varepsilon\,|\,(ab)^+\\
&= a(a\,|\,b)^\star b\,|\,(ab)^+\,|\,(ba)^\star && \textit{since}\, \{\varepsilon\} \subset L((ba)^\star)
\end{aligned}$$

We have:

$$\begin{aligned}
(ab)^+ &= (ab)(ab)\dots(ab)\\
&= a(ba)(ba)\dots(ba)b\,|\,ab\\
&= a(ba)^\star b
\end{aligned}$$

Also $(ba) \subset (a\,|\,b)^\star$ and then $(ba)^\star \subset (a\,|\,b)^\star$, because $(a\,|\,b)^\star$ contains all the words. Therefore $a(ba)^\star b \subset a(a\,|\,b)^\star b$, i.e. $(ab)^+ \subset a(a\,|\,b)^\star b$.

As a consequence, one possible answer is

$$a(a\,|\,b)^\star b\,|\,(ab)^\star\,|\,(ba)^\star = a(a\,|\,b)^\star b\,|\,(ba)^\star$$

The intersection between $L(a(a\,|\,b)^\star b)$ and $L((ba)^\star)$ is empty because all the words of the former start with a, while all the words of the other start with b (or is ε). Therefore we cannot simply further this way.