Fundamentele Informatica II

Answer to selected exercises 1

John C Martin: Introduction to Languages and the Theory of Computation

M.M. Bonsangue (and J. Kleijn)

Fall 2011

• *Let L be a language. It is clear from the definition that $L^+ \subseteq L^*$. Under which circumstances are they equal?*
By definition $L^* = \bigcup_{i\geq 0} L^i = L^0 \cup \bigcup_{i\geq 1} L^i = \{\Lambda\} \cup L^+$ for any language L. Hence, it is clear that $L^+ \subseteq L^*$. Moreover, we see immediately that $L^* = L^+$ if and only if $\Lambda \in L^+$. We claim that $\Lambda \in L^+$ if and only if $\Lambda \in L$. This can be proved as follows:
Clearly, if $\Lambda \in L$, then $\Lambda \in L^+$.
Now assume that $\Lambda \in L^+$. Thus there exists an $i \geq 1$ such that $\Lambda \in L^i$. Hence Λ is the shortest word in L^i. Since the length of any shortest word in L is i times the length of a shortest word in L, it follows that $\Lambda \in L$.
The claim being proved, we have $L^* = L^+$ if and only if $\Lambda \in L$.

• *Find a language L over $\{a, b\}$ that is neither $\{\Lambda\}$ nor $\{a, b\}^*$ and satisfies $L = L^*$.*
First of all observe that $L = L^*$ implies that $\Lambda \in L$. Moreover, in combination with $L \neq \{\Lambda\}$ it follows that L must be an infinite language.
A good example is $\{a\}^*$, because $\{a\}^* \neq \{\Lambda\}$ and $(\{a\}^*)^* = \{a\}^*$, since $\{a\}^*\{a\}^* = \{a\}^*$.
Another example is $\{x \in \{a, b\}^* \mid |x| \text{ is even}\}$.

• *Find an infinite language L over $\{$ a,b$\}$ for which $L \neq L^*$.*
Observe that each *finite* language such that $L \neq \{\Lambda\}$ has the property that $L \neq L^*$, even $\emptyset^* = \{\Lambda\} \neq \emptyset$. This follows from the fact that $\Lambda \in L^*$ whatever L. Similarly: for very infinite language L such that $\Lambda \notin L$, we have $L \neq L^*$. Consequently, example languages as requested are, e.g., $\{a\}^+$, $\{a, b\}^+$, and $\{x \in \{a, b\}^* \mid |x| \text{ is odd}\}$; observe that $\{x \in \{a, b\}^* \mid |x| \text{ is odd}\}^*$ contains not only Λ, but actually all even words: $\{x \in \{a, b\}^* \mid |x| \text{ is odd}\}^* = \{a, b\}^*$.

• *Give examples of languages* L_1 *and* L_2 *such that* $L_1L_2 = L_2L_1$ *and*
a. $L_1 \neq \{\Lambda\} \neq L_2$ *and neither language contains the other one:*
Take $L_1 = \{a\}$, and $L_2 = \{aa\}$.
b. $\emptyset \neq L_1 \subset L_2$ *and* $L_1 \neq \{\Lambda\}$:
Take $L_1 = \{a\}$ and $L_2 = \{\Lambda, a\}$.

• *Show that for any language* L, $L^* = (L^*)^* = (L^+)^* = (L^*)^+$.
Whatever the language L, it always holds that $L \subseteq L^+ \subseteq L^*$, by the definition of $+$ and $*$.
Then exercise 1.33 implies that $L^* \subseteq (L^+)^* \subseteq (L^*)^*$.
Moreover, $L^* \subseteq (L^*)^+ \subseteq (L^*)^*$.
Now assume that the inclusion $(L^*)^* \subseteq L^*$ is always true. Then $L^* = (L^*)^*$ and all inclusions above are equalities: $L^* = (L^+)^* = (L^*)^+ = (L^*)^*$.
Thus the only thing left to prove is that the inclusion $(L^*)^* \subseteq L^*$ is always true. Consider $w \in (L^*)^*$. Thus there exist a $k \geq 0$ and words $v_1, \ldots, v_k \in L^*$ such that $w = v_1 \cdots v_k$ (note that $w = \Lambda$ if $k = 0$). Thus w is a concatenation of 0 or more words from L, in other words: $w \in L^*$.

1.32 If L is a finite set, then $|L|$ denotes the number of elements (the cardinality) of L.
Let L_1 and L_2 be two finite languages. Then, clearly, $|L_1L_2| \leq |L_1||L_2|$ because every element of L_1L_2 is obtained by combining an element from L_1 with one from L_2 and there are $|L_1| \times |L_2|$ ways to do this. It is however not necessarily the case that $|L_1L_2| = |L_1||L_2|$, because different choices may still yield the same result. Let $L_1 = L_2 = \{a, a^2\}$ then $L_1L_2 = \{aa, aaa, aaaa\}$ and $|L_1L_2| = 3 \neq 4$.
Another example are $L_1 = \{a, ab\}$, $L_2 = \{a, ba\}$. Then $L_1L_2 = \{aa, aba, abba\}$.

1.33 Let $L_1, L_2 \subseteq \{a, b\}^*$. **a.** Assume $L_1 \subseteq L_2$. Then $L_1^2 = L_1L_1 \subseteq L_1L_2 \subseteq L_2^2$ and in general $L_1^k = L_1^{k-1}L_1 \subseteq L_1^{k-1}L_2 \subseteq L_2^{k-1}L_2 = L_2^k$ for all $k \geq 1$. Consequently,
$L_1^* = \{\Lambda\} \cup L_1 \cup L_1^2 \cup \ldots \cup L_1^k \cup \ldots \subseteq \{\Lambda\} \cup L_2 \cup L_2^2 \cup \ldots \cup L_2^k \cup \ldots = L_2^*$.
b. $L_1^* \cup L_2^* \subseteq (L_1 \cup L_2)^*$ always holds, since $L_1 \subseteq L_1 \cup L_2$ which implies that $L_1^* \subseteq (L_1 \cup L_2)^*$ (see item a. above) and similarly $L_2^* \subseteq (L_1 \cup L_2)^*$.
c. The inclusion $L_1^* \cup L_2^* \subseteq (L_1 \cup L_2)^*$ may be strict: for $L_1 = \{0\}$ and $L_2 = \{1\}$ we have $\{0\}^* \cup \{1\}^* \neq \{0, 1\}^*$.
d. If $L_1^* \subseteq L_2^*$ then $L_1^* \cup L_2^* = L_2^*$ and since $L_1 \subseteq L_1^* \subseteq L_2^*$ also $(L_1 \cup L_2)^* \subseteq (L_2^* \cup L_2)^* = L_2^*$. Similarly, $L_2^* \subseteq L_1^*$ implies that $L_1^* \cup L_2^* = L_1^* = (L_1 \cup L_2)^*$.
Next consider $L_1 = \{0^2, 0^5\}$ and $L_2 = \{0^3, 0^5\}$.
Then $L_1^* = \{\Lambda, 0^2, 0^4, 0^5, \ldots\} = \{0\}^* - \{0, 0^3\}$ and

$L_2^* = \{\Lambda, 0^3, 0^5, 0^6, 0^8, 0^9, \ldots\} = \{0\}^* - \{0, 0^2, 0^4, 0^7\}$. Thus neither $L_1^* \subseteq L_2^*$ nor $L_2^* \subseteq L_1^*$. However, $L_1^* \cup L_2^* = \{0\}^* - \{0\}$ and also $(L_1 \cup L_2)^* = \{0^2, 0^3, 0^5\}^* = \{\Lambda, 0^2, 0^3, 0^4, 0^5, \ldots\} = \{0\}^* - \{0\}$.

• *List some elements of $\{a, ab\}^*$ and give a proposition describing all and only these elements. Further try to find a procedure to test if a word x satisfies your proposition.*
$\{a, ab\}^*$ contains (among others) the following words: Λ, a, ab, aab, aba, aa, $abab$, aaa, $aaab$, $aaba$, etc.
a. $\{a, ab\}^*$ is precisely the set of strings in which every b is preceded by at least one a or —alternatively— $\{a, ab\}^*$ is precisely the set of strings which do not start with b and do not have a subword bb.
b. Hence a procedure to test whether a word belongs to $\{a, ab\}^*$ is to simply go from left to right through the string, symbol by symbol: it should not begin with b and after every occurrence of b either the next letter is an a or the end of the string has been reached.

1.36 L consists of all strings from $\{a, b\}^*$ that do not end with b and do not have a subword bb.
a. $L = \{a, ba\}^*$.
b. Consider now the language K consisting of all strings from $\{a, b\}^*$ that do not have a subword bb. Assume that $K = S^*$ for a finite set S. Then $b \in K = S^*$. Since $S^* = S^*S^*$, it follows that $bb \in S^*S^* = S^* = K$, a contradiction. Hence there cannot exist a finite S such that $K = S^*$.

1.37 Let $L_1, L_2, L_3 \subseteq \Sigma^*$ for some alphabet Σ.
a. $L_1(L_2 \cap L_3) \subseteq L_1L_2 \cap L_1L_3$, because $w \in L_1(L_2 \cap L_3)$ implies that $w = xy$ with $x \in L_1$ and $y \in L_2 \cap L_3$. Consequently, $w \in L_1L_2$ and $w \in L_1L_3$.
Equality does not necessarily hold. Let $L_1 = \{a, ab\}$, $L_2 = \{ba\}$, and $L_3 = \{a\}$. Then $L_1(L_2 \cap L_3) = \emptyset \neq \{aba\} = \{aba, abba\} \cap \{aa, aba\} = L_1L_2 \cap L_1L_3$,
b. $L_1^* \cap L_2^* \supseteq (L_1 \cap L_2)^*$, because $w \in (L_1 \cap L_2)^*$ implies that w is a concatenation of 0 or more words from $L_1 \cap L_2$. Consequently, $w \in L_1^*$ and $w \in L_2^*$.
Equality does not necessarily hold. Let $L_1 = \{a\}$ and $L_2 = \{aa\}$. Then $L_1^* \cap L_2^* = \{a\}^* \cap \{aa\}^* = \{aa\}^* \neq \{\Lambda\} = \emptyset^* = (L_1 \cap L_2)^*$.
c. $L_1^*L_2^*$ and $(L_1L_2)^*$ are not necessarily included in one another.
Let $L_1 = \{a\}$ and $L_2 = \{b\}$. Then $L_1^*L_2^* = \{a\}^*\{b\}^*$ consisting of words with a number of a's followed by some number of b's and $(L_1L_2)^* = \{ab\}^*$ consisting of words with alternating a's and b's. These two languages are incomparable: $aab \in L_1^*L_2^* - (L_1L_2)^*$ and $abab \in (L_1L_2)^* - L_1^*L_2^*$.

• *Let $x, y \in \Sigma^*$ for some alphabet Σ. Whereas in general xy and yx are two different words, equality is possible, for instance if $x = \Lambda$ or $y = \Lambda$. Can this still happen if x and y are both nonnull? Describe the precise conditions when this can happen.*
Assume that $x \neq \Lambda$ and $y \neq \Lambda$ and $xy = yx$ holds. Before giving a characterization of the conditions allowing this situation, we first informally explore what is going on.
If $|x| = |y|$, then it must be the case that $x = y$.
If $|x| \neq |y|$ we may assume that $|x| > |y|$; the other case follows by symmetry. $|x| > |y|$ in combination with $xy = yx$ implies that there exists a non-empty word z such that $x = yz = zy$. Since $y \neq \Lambda$, we have $|yz| = |x| < |yx|$ and we may use induction to prove the following
Claim For all $s, t \in \Sigma^*$ such that $s \neq \Lambda \neq t$: $st = ts$ if and only if there exists a word u and natural numbers p, q such that $s = u^p$ and $t = u^q$.
*Proof of claim*The if-direction is obvious: $st = u^p u^q = u^{p+q} = u^q u^p = ts$.
For the only-if-direction we use induction on $|st|$:
If $|s| = 1 = |t|$, then $s = a$ and $t = b$ for some $a, b \in \Sigma$. From $st = st$ it follows that $a = b$. Hence we let $u = a$ and $p = q = 1$ and we are done.
Next assume that $|s| > |t| \geq 1$. Then as argued above, there exists a word z such that $s = tz = zt$. Since $|tz| = |tz| < |st| = |ts|$ we can apply the induction hypothesis: there exists a word u and natural numbers r, q such that $z = u^r$ and $t = u^q$. Consequently, $s = u^{r+q}$ and we are done with $p = r + q$.

• *Show that there is no language L so that $L^* = \{aa, bb\}^*\{ab, ba\}^*$.* We prove by contradiction that exists no language L such that $L^* = \{aa, bb\}^*\{ab, ba\}^*$. Suppose that L is such that $L^* = \{aa, bb\}^*\{ab, ba\}^*$. Consequently, $aa \in L^*$ and $ab \in L^*$ which implies that $abaa \in L^*$. However, $abaa$ is not an element of $\{aa, bb\}^*\{ab, ba\}^*$.

• *Let $L = \{x \in \{0,1\}^* \mid x = yy$ for some string $y\}$. Prove or disprove that there exist two languages $L_1 \neq \{\Lambda\}$ and $L_2 \neq \{\Lambda\}$ such that $L_1 L_2 = L$.*
We prove by contradiction that the statement is not true:
Suppose L_1 and L_2 are such that $L_1 L_2 = L$. Observe that $\{00\}^*$ and $\{11\}^*$ are both subsets of $L = L_1 L_2$.
If $0^i \in L_1$ and $1^j \in L_2$ for some $i, j \geq 1$, then $0^i 1^j \in L_1 L_2 = L$, a contradiction. Similarly, we arrive at a contradiction if $1^i \in L_1$ and $0^j \in L_2$ for some $i, j \geq 1$.
Hence it must be the case that either L_1 or L_2 contains only "mixed" strings with at least one occurrence of a 0 and at least one occurrence of a 1. We

only consider the case that L_1 has this property. The case that it would be L_2 is symmetrical.
Let $w = x_1 0 x_2 1 x_3$ be such a word. Consider $1^j \in L_2$ with $j \geq |w|$. Then $x_1 0 x_2 1 x_3 1^j \in L_1 L_2$, but it cannot be an element of L, a contradiction.
The remaining case ($w = x_1 1 x_2 0 x_3$) is dealt with in an analogous way. Hence we arrive in all cases at a contradiction and the assumed languaes L_1, L_2 cannot exist.

---

version 2 September 2011