CS 173 (B), Spring 2015, Examlet 3, Part A

NAME:	**NETID:**

Discussion Section: BDA:1PM BDB:2PM BDC:3PM BDD:4PM BDE:5PM

The following problems relate to the grammar below. S is the start symbol, and a, b are the terminals.

$$S \to aS \mid aX \mid XS$$
$$X \to \epsilon \mid XaXbX \mid XbXaX$$

Before attempting the problems below, you may want to try and understand the strings generated by the variable X in this grammar.

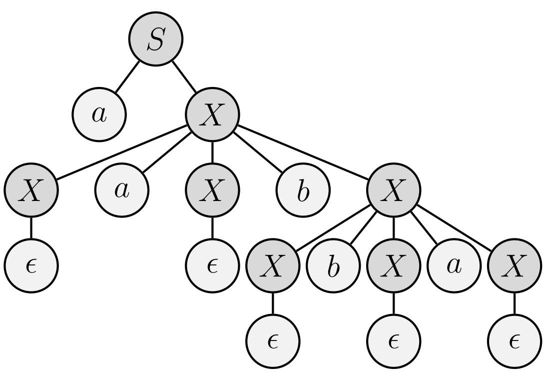
1. **Parse Tree** [10 points]

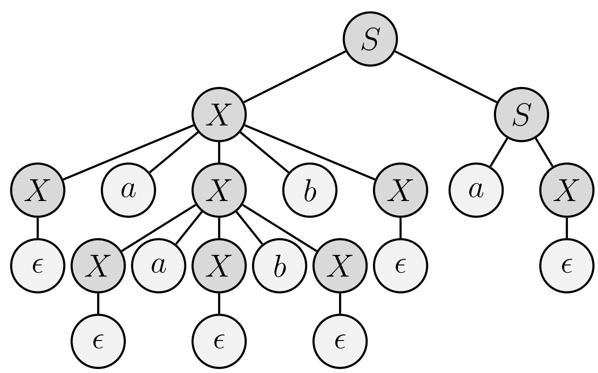
   Give two different parse trees for the string $aabba$ according to the above grammar.

   Solution:

   There are infinitely many parse trees for this string. The most "natural" ones are shown below.

   It maybe helpful to first notice that all the strings generated by X have equal number of a's and b's. Hence we could use the rule $S \to aX$ (with X generating $abba$) or the rule $S \to XS$ (with X generating $aabb$) as the top-level rule.





   It is also possible to use $S \to XS \to \epsilon S$ as the first rule, and generate the whole string from S again (using either of the above two parse trees).

   However, the rule $S \to aS$ cannot appear in the generation of this string.

2. **Proof by Induction** [15 points]

Prove by induction that this grammar generates only strings with more a's than b's.

[Hint: Prove something about the strings generated by the variable X, in addition to the claim in the problem.]

Solution: We prove the following stronger claim.

Claim: Strings of terminals generated by parse-trees with X at the root have equal number of a's and b's, and strings of terminals generated by parse-trees with S at the root have more a's than b's.

Proof: We prove this by strong induction on the height of the parse-trees, for all parse-trees of height ≥ 1 (since there are no parse trees of height zero which generate a string of terminals).

Base case: There is exactly one parse-tree of height $= 1$ with X at the root that generates a string of terminals: it generates the string ϵ, which has equal number of a's and b's. The second part of the claim is vacuously true for height $= 1$, since there are no parse-trees of height 1 with S at the root.

Induction Hypothesis: Suppose, for some $k \geq 1$, all strings of terminals generated by parse-trees of height $h \leq k$ with X at the root have equal number of a's and b's; and all strings of terminals generated by parse-trees of height $h \leq k$ with S at the root have more a's than b's.

We claim that the same properties hold for strings of terminals generated by parse-trees of height $k+1$, with X and S at the root, respectively.

Consider any string of terminals α, generated by a parse-tree of height $k+1$, with X at the root. In this parse tree, the root node must be expanded using one of the rules $X \to XaXbX$ or $X \to XbXaX$ (since $k+1 > 1$, the root cannot be expanded using the rule $X \to \epsilon$). In either case, the three subtrees rooted at the children of the root, labeled by X have heights $\leq k$, and so by the induction hypothesis, each one of them generates a string with equal number of a's and b's. Let n_1, n_2, n_3 be the number of a's (and hence number of b's) in these three strings. Then the string α has $n_1 + n_2 + n_3 + 1$ a's and the same number of b's, proving the first part of the claim.

Next consider any string of terminals α, generated by a parse-tree of height $k+1$, with S at the root. In this parse tree, the root node must be expanded using one of the rules $S \to aS$, $S \to aX$ and $S \to XS$. Let ℓ_a, ℓ_b be the number of a's and b's respectively, in the string generated by the left child of the root, and n_a, n_b be those numbers in the string generated by the right child of the root. Then $\ell_a + n_a$ and $\ell_b + n_b$ are the number of a's and b's in α. In the first case, $\ell_a = 1$, $\ell_b = 0$ and $n_a > n_b$. In the second case, $\ell_a = 1$, $\ell_b = 0$, $n_a = n_b$. In the third case, $\ell_a = \ell_b$ and $n_a > n_b$. Hence in all three cases, $\ell_a + n_a > \ell_b + n_b$. Thus the second part of the claim holds as well.

CS 173 (B), Spring 2015, Examlet 3, Part A

NAME:	**NETID:**

Discussion Section: BDA:1PM BDB:2PM BDC:3PM BDD:4PM BDE:5PM

The following problems relate to the grammar below. S is the start symbol, and a, b are the terminals.

$$S \to Sa \mid Xa \mid SX$$
$$X \to \epsilon \mid aXb \mid bXa \mid XX$$

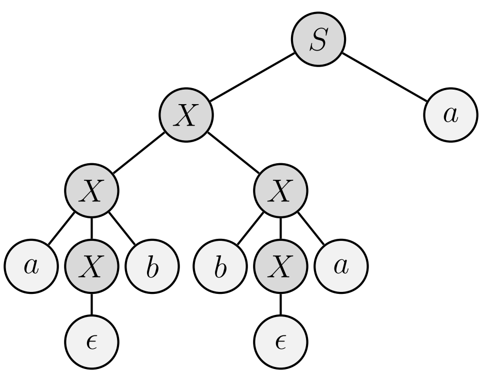
Before attempting the problems below, you may want to try and understand the strings generated by the variable X in this grammar.

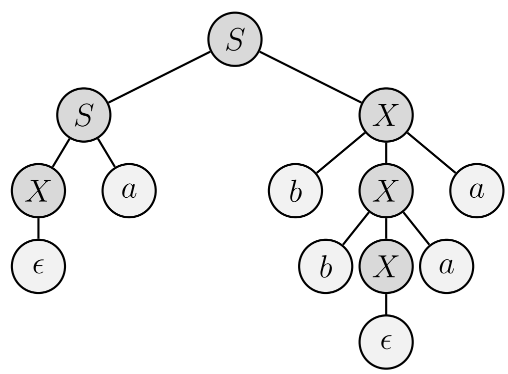
1. **Parse Tree** [10 points]

   Give two different parse trees for the string $abbaa$ according to the above grammar.

   There are infinitely many parse trees for this string. The most "natural" ones are shown below.

   It maybe helpful to first notice that all the strings generated by X have equal number of a's and b's. Hence we could use the rule $S \to Xa$ (with X generating $abba$) or the rule $S \to SX$ (with X generating $bbaa$) as the top-level rule.





   It is also possible to use $S \to XS \to \epsilon S$ as the first rule, and generate the whole string from S again (using either of the above two parse trees).

   However, the rule $S \to aS$ cannot appear in the generation of this string.

2. **Proof by Induction** [15 points]

Prove by induction that this grammar generates only strings with more a's than b's.

[Hint: Prove something about the strings generated by the variable X, in addition to the claim in the problem.]

Solution: We prove the following stronger claim.

Claim: Strings of terminals generated by parse-trees with X at the root have equal number of a's and b's, and strings of terminals generated by parse-trees with S at the root have more a's than b's.

Proof: We prove this by strong induction on the height of the parse-trees, for all parse-trees of height ≥ 1 (since there are no parse trees of height zero which generate a string of terminals).

Base case: There is exactly one parse-tree of height $= 1$ with X at the root that generates a string of terminals: it generates the string ϵ, which has equal number of a's and b's. The second part of the claim is vacuously true for height $= 1$, since there are no parse-trees of height 1 with S at the root.

Induction Hypothesis: Suppose, for some $k \geq 1$, all strings of terminals generated by parse-trees of height $h \leq k$ with X at the root have equal number of a's and b's; and all strings of terminals generated by parse-trees of height $h \leq k$ with S at the root have more a's than b's.

We claim that the same properties hold for strings of terminals generated by parse-trees of height $k+1$, with X and S at the root, respectively.

Consider any string of terminals α, generated by a parse-tree of height $k+1$, with X at the root. In this parse tree, the root node must be expanded using one of the rules $X \to aXb$, $X \to bXa$ or $X \to XX$ (since $k+1 > 1$, the root cannot be expanded using the rule $X \to \epsilon$). In all cases, the subtree rooted at each child of the root labeled by X has height $\leq k$, and so by the induction hypothesis, it generates a string with equal number of a's and b's. In the first two cases, let n be the number of a's (and hence number of b's) in this string; then the string α has $n+1$ a's and the same number of b's. In the third case, let n_1, n_2 be the number of a's (and hence number of b's) in the two strings generated by the parse-trees rooted at the two children of the root; then the original string has $n_1 + n_2$ a's and b's. Thus in all cases, the first part of the claim holds.

Next consider any string of terminals α, generated by a parse-tree of height $k+1$, with S at the root. In this parse tree, the root node must be expanded using one of the rules $S \to Sa$, $S \to Xa$ and $S \to SX$. In all cases, the subtrees rooted at the children of the root node have height $\leq k$, and the induction hypothesis applies to them. Let n_a, n_b be the number of a's and b's respectively, in the string generated by the left child of the root, and r_a, r_b be those numbers in the string generated by the right child of the root. Then $n_a + r_a$ and $n_b + r_b$ are the number of a's and b's in α. In the first case, $n_a > n_b$, $r_a = 1$, $r_b = 0$. In the second case, $n_a = n_b$, $r_a = 1$, $r_b = 0$. In the third case, $n_a > n_b$, $r_a = r_b$. Hence in all three cases, $n_a + r_a > n_b + r_b$. Thus the second part of the claim holds as well.