String algebra: in principle

Suppose that Σ is a finite "***alphabet***" (set) of ***characters*** or ***symbols***.

1. Define what we mean by a ***string*** (or ***word***) over the alphabet Σ.
   (a) Compare the abstract concept of a string with those used in programs.
   (b) What do we mean by the string ε?
   (c) Given a string x, what does $|x|$ represent? What is $|\varepsilon|$?

2. If x and y are strings over Σ, define their ***product*** xy, i.e., $x \cdot y$.
   (a) What other term describes this operation?
   (b) What properties does this operation have, and—crucially—what key property does it *not* have?
   (c) How does the product let us define nonnegative whole ***powers*** z^n of a string z?
   (d) If x and y are strings, how can we simplify $|xy|$ and $|x^n|$?

3. If W is a set of strings, define its ***asterate*** W^*.
   (a) What properties does this operation have, and what string does this set *always* include?
   (b) If x is a single character or string over Σ, we can define $x^* = \{x\}^*$.
   What strings does the set x^* contain?
   (c) In words, what is Σ^*, and what does the statement "$x \in \Sigma^*$" mean?

4. Blurring the line between strings and sets of strings a bit, if x and y are strings, what is meant by their formal ***sum*** $x + y$, and what set operation does this correpond to?
   What properties does this operation have, and how does it relate to our other operation "$\cdot$"?

5. What is a ***regular expression*** over Σ?

. . . and in practice

In the problems below, characters in Σ will be written in `typewriter font`, with variables representing strings written as usual via *italic letters*.

6. Constrast, in writing, the meanings of the following,
   including the context of whether each is an *element* of Σ^* or a *subset* of Σ^*: (a) ε (b) $\varnothing$ (c) $\{\varepsilon\}$

7. For any set W of strings, describe in writing what $W^* \setminus \{\varepsilon\}$ means.
   [This comes up enough that we often denote it by W^+.]

8. Simplify the following string expressions:
   (a) $\mathtt{a}^3\varepsilon^5(\mathtt{bca})^2$ (b) $\mathtt{a}(\varepsilon + \mathtt{b} + \mathtt{c})w$ (c) $(\varepsilon + \mathtt{a})(\varepsilon + \mathtt{b})$ (d) $(\mathtt{a} + xy)^3$

9. Describe the sets of strings generated by the following regular expressions over $\Sigma = \{0, 1\}$, both in writing and using set-builder notation:
   (a) $\mathtt{1}^*$ (b) $\mathtt{01}^*$ (c) $\mathtt{0}^*\mathtt{1}$ (d) $(\mathtt{01})^*$ (e) $\mathtt{0}^*\mathtt{1}^*$

10. Find regular expressions expressing the following sets of strings over $\Sigma = \{0, 1\}$:
    (a) all strings starting with `101`;
    (b) all strings containing the substring `000`;
    (c) all strings that can be built as a products of some number of copies of the string `001`; and
    (d) all strings in which a `0` is not preceded by a `1`.