Strings: Basic notation (Solutions)

EXERCISE 1.
Fill in = or $\neq$ as appropriate for each pair of strings below.

1. $abba$ _ $ABBA$
2. 10 _ $5 + 5$
3. $\{m, a, d\}$ _ $\{d, a, m\}$

Caution: { and } can be symbols just like m, a, or d.
Solution

1. $abba \neq ABBA$
2. $10 \neq 5 + 5$
3. $\{m, a, d\} \neq \{d, a, m\}$

Explanation

1. The symbols a and A are distinct (and so are b and B). Hence $abba$ and $ABBA$ already contain different symbols in their very first position, and consequently the strings are distinct.
2. While it is true that the sum of 5 and 5 is 10, the exercise asks you to compare the string $5+5$ to the string 10. These are very different strings. For example, only the former contains the symbol +, and only the letter contains 1 and 0.
3. This one is tricky. If we were talking about sets, then it would be the case $\{m, a, d\}$ is equivalent to $\{d, a, m\}$ because sets are unordered. But here you are asked to consider strings, not sets. These strings just so happen to start with { and end with }, but they are still strings. And since the second symbol of the left string is m, whereas the right string has d in its second position, these two strings are distinct.

EXERCISE 2.
For each one of the following, say whether it is a valid alphabet. Justify your answer.

- $\{a\}$
- $\{0, 1\}$
- the set of all English words that are spelled with at most 5 characters
- all natural numbers less than 1000
- the nucleobases of DNA: adenine, cytosine, guanine, thymine

Solution
An alphabet must be a finite set of symbols. Therefore:

- $\{a\}$ is valid because it contains exactly one symbol and thus is finite
- $\{0, 1\}$ is valid because it contains exactly two symbols and thus is finite
- The set of all English words that are spelled with at most 5 characters is also a finite set: The English alphabet has 26 letters, which means that there are at most $26^5 = 11,881,376$ words that fit these requirements. While that is a large number, it is still finite. The fact that each symbol is itself a string of characters does not matter, that is an artefact of our writing system. If instead we had a unique symbol to refer to each one of these words, e.g. pictograms, the issue wouldn't even arise.
- The set of all natural numbers less than 1000 is also a possible alphabet because this set is finite. Again you might be concerned that we write natural numbers as strings of digits and it may seem weird to you to have symbols that are internally complex. But just like with the English words, this is just an artefact of how we write numbers. In hexademical notation, for instance, the natural number 11 would be a single symbol b.
- The nucleobases of DNA once again form a finite set and hence are a valid alphabet. Again these are symbols with internal structure, in this case their molecular structure. When we treat the nucleobases as the symbols of our alphabet, we are making a decision to ignore their internal structure and treat them as the smallest unit of analysis (just like, say, a phonologist may treat sounds as the smallest unit of analysis and ignore the sound waves that actually give rise to the sounds).

EXERCISE 3.
Which one of the following are members of $\{a, b\}^4$, i.e. Σ^4 where Σ contains a, b, and nothing else?

1. *aaab*
2. *aba*
3. *aaaaa*
4. *b*
5. *abca*

Solution
Among the listed strings, $aaab$ is the only member of $\{a, b\}^4$.
Explanation
The $\{a, b\}^4$ contains all strings that meet both of the following criteria:

1. The only symbols that may occur in the string are a and b.
2. The string must contain exactly four symbols.

The only strings above with exactly four symbols are $aaab$ and $abca$, and only the former contains no symbols besides a and b. Note that $aaaa$, which isn't listed

above, is also a member $\{a, b\}^4$ — the requirement that the strings may contain no symbols other than a and b does not entail that the strings must contain both a and b.

EXERCISE 4.
List all members of $\{k, o, z\}^2$.

Solution
The members are *kk, ko, kz, ok, oo, oz, zk, zo, zz*.

Explanation
The set $\{k, o, z\}^2$ contains all strings of length 2 that can be built using only the symbols k, o, and z. This gives us three options starting with k (*kk, ko, kz*), three options starting with o (*ok, oo, oz*), and three options starting with z (*zk, zo, zz*).

EXERCISE 5.
Write each one of the following in a more compact fashion using exponents.

1. ABBA
2. loool
3. aardvark

Solution

1. AB^2A
2. lo^3l
3. $a^2rdvark$

Explanation
The general strategy is to find parts of the string where the same symbol x is iterated n times and to then replace that part with x^n. For example, in *ABBA* we have A followed by two Bs and then another A, i.e. AB^2A.

EXERCISE 6.
Enumerate the five shortest members of $\{a\}^*$.

Solution

1. ε,
2. a,
3. aa,
4. aaa,
5. $aaaa$

Explanation

3

The expression $\{a\}^*$ is short for "the set of all strings of length 0 or more that can be built over the alphabet $\{a\}$". With only one symbol to choose from, there can never be two distinct strings of the same length. Hence we just start with the shortest possible string, which is the empty string ε, and then keep adding as until we have enumerated five strings.

EXERCISE 7.

Give an example of distinct u and v such that $u \cdot v = v \cdot u$ and neither u nor v is the empty string.

Solution

There are infinitely many solutions. For example, if $u := aba$ and $v := abaaba$, then $u \cdot v = aba \cdot abaaba = abaabaaba = abaaba \cdot aba = v \cdot u$. The simplest solution is actually $u := a$ and $v := aa$, where we have $u \cdot v = a \cdot aa = aaa = aa \cdot a = v \cdot u$.

Note that $u := a$ and $v := ab$ is not a valid solution. In this case, $u \cdot v = aab$, whereas $v \cdot u = aba$, which is distinct from aab. Similarly, $u := a$ and $v := a$ is not a valid solution because the exercise requires u and v to be distinct strings.

EXERCISE 8.

Is the following true or false? If $u \neq v$, then $u \cdot v \neq v \cdot u$?

Solution

False.

Explanation

This follows immediately from the previous exercise, where you had to pick distinct u and v such that $u \cdot v \neq v \cdot u$.