Solution to Problem 13

Task. Show that the language L of all the words that have equal number of u's and d's and twice as many s's is not context-free.

Solution. Let us prove that this language is not context-free.

The proof will be by contradiction. Let us assume that this language is context-free. Then, by pumping lemma, there exists an integer p such that every word w from the language L whose length is at least p can be represented as $w = uvxyz$, where:

- $\text{len}(vxy) \le p$;
- $\text{len}(vy) > 0$, and
- for all natural numbers i, the word $uv^i xy^i z$ also belongs to the language L.

Let us take the word

$$w = u^p d^p s^{2p} = u \ldots u d \ldots d s \ldots s,$$

where first u is repeated p times, then d is repeated p times, and finally s is repeated $2p$ times. The length of this word – i.e., the number of symbols in this word – is equal to $p+p+2p = 4p$. Clearly, $4p \ge p$, so, according to the Pumping Lemma, this word can be described as $uvxyz$ with the above properties.

Where can the central part vxy of this word be? We know that the length $\text{len}(vxy)$ of this part cannot exceed p. Thus, it cannot contain three different types of symbols: u's, d's, and s's – since then it would have to include all p symbols d plus additional u and s symbols, so its length would have been larger than p. So, there are only 5 cases remaining for the location of the part vxy:

1. it can be in the u's;
2. it can be in u's and d's;
3. it can be in d's;
4. it can be in d's and s's;
5. it can be in s's.

Let us consider these cases one by one.

Case 1. If vxy is in the u's, this means that the parts v and y contain only u's. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add u's – but we do not add any d's or s's. In the original word $w = u^pd^ps^{2p}$, there was a balance between letters of the three types. When we add more u's, the balance is disrupted: the ratio between number of u's and number of d's is now different from 1. Since the language L only contains the words for which the above proportions have to be satisfied, the word $uvvxyyz$ cannot belong to the language L.

Case 2. If vxy is in u's and in d's, this means that the parts v and y contain only u's and d's. Thus, when we pump, i.e., when we go from the original word $uvxyz$ to the word $uv^2xy^2z = uvvxyyz$, we add u's and/or d's – but we do not add any s's. In the original word $w = u^pd^ps^{2p}$, there was the desired balance between numbers of letters of all three types. When we add more u's and/or d's, the balance is disrupted, so the word $uvvxyyz$ cannot belong to the language L.

Similarly, we can see that in the other 3 cases, we also get a contradiction. This means that the original assumption – that the language L is context-free – is wrong. Thus, the language L is not context-free.