Input: a text from a specified language.

Steps:

1. Change all letters to lowercase.
2. Remove all types of dashes and hyphens (i.e. in Croatian one can write crno-beli, the text after applying this rule will be crnobeli).
3. Delete all words which contain anything (letters, numbers, special signs) which do not belong to an official alphabet (specified for all languages).
4. For each language, a list of words which do not contain a vowel is provided (mostly prepositions, like k, s, z, v in Slovak). Delete all other words which do not contain a vowel.
5. Attach all allowed words without a vowel to the preceding/following words, according to rules provided for each language.
6. Syllabification itself:
7. Preliminarily, each syllable ends after a vowel or after a syllabic consonant (a list of syllabic consonants, if there are any, is provided for each language).
8. If there remains a consonant or a consonant cluster at the end of a word, attach it to the last consonant.
9. Check whether the sonority principle is kept in each syllable except for the first syllable in each word.

According to the sonority principle, the sonority of a syllable is not allowed to decrease from the beginning to the vowel/syllabic consonants. Sonority is defined as follows: vowels and syllabic consonants – sonority value 4; glides - sonority value 3, liquids - sonority value 2, nasals – sonority value 1; all other consonants (obstruents) – sonority value 0.

If the sonority principle is violated (possible only if one (or more) sonorant precedes a non-sonorant consonant before a vowel/syllabic consonant), attach the “problematic sonorant” to the end of the previous syllable.

It is allowed “not to obey” the sonority principle in the first syllable of a word (and, consequently, in all one-syllabic words), e.g. “lži” or “lživý” in Slovak.

Examples from Croatian:

derivacija; after a) we have de-ri-va-ci-ja, everything ok

integral; after a) we have i-nte-gra-l, applying b) we obtain i-nte-gral; now the syllable nte does not respect the sonority hierarchy, as the sequence of sonority values in the syllable is 1 0 4, so we move the letter n to the previous syllable, finally obtaining in-te-gral.

Output:

* a syllabified text,
* an excel table with the list of all syllables which occur in the text (column 2), their frequencies in the text (column 3), and their length measured in the number of phonemes (the rules are provided for each language), the syllables will be order according to their frequency, with the most frequent syllable being the first, etc.,
* an excel table with the frequency table of lengths of syllables, with all occurring syllables taken into account (tokens – i.e., if the syllable “ma” with length 2 occurs 500-times in the text, it contributes to the frequency of length 2 500-times),
* an excel table with the frequency table of lengths of syllables, only with different syllables taken into account (types – i.e., if the syllable “ma” with length 2 occurs 500-times in the text, it contributes to the frequency of length 2 only once).