
ГЕНЕРИРАНЕ НА ИЗРЕЧЕНИЯ

Да се напише програма, която по зададено множество думи и техните характеристики строи на произволен принцип „смислени“ изречения от тях.

Нека за този проект приемем, че думите в естествения ни език се делят само на четири вида: **определения**, **подлози**, **сказуеми** и **допълнения**. Тогава на изречението “Голямото куче лае птиците.” ще правим следния разбор: думата “**голямото**” е определение, “**куче**” е подлог, “**лае**” е сказуемо и “**птиците**” е допълнение. Ще приемаме, че други видове думи за нас не съществуват. Всеки вид дума ще носи със себе си най-много два етикета:

- Етикет, указващ каква роля може да играе думата в изречението – точно един измежду стринговете “**о**”, “**п**”, “**с**” и “**д**” за съответните видове думи.
- Етикет, указващ какъв род или какво число е думата – за подлозите и определенията им трябва да бъде точно един измежду стринговете “**м**”, “**ж**”, “**ср**” и “**мн**”. За сказуемите трябва да бъде само един измежду “**ед**” и “**мн**”, тъй като за тях родът на подлога няма значение, само числото. Допълненията нямат свойства, които да ни интересуват, следователно не носят такъв етикет въобще.

Допълненията могат дори да бъдат по-сложни изрази, като в изречението “Зелените жаби гледат **към небето**.”. Думите от тези примерни изречения ще бъдат представени на езика Scheme в базата данни по следния начин (подредбата им няма значение):

- (“голямото” “о” “ср”) – определение, среден род (следователно единствено число)
- (“куче” “п” “ср”) – подлог, среден род (също единствено число)
- (“лае” “с” “ед”) – сказуемо, единствено число – за него родът няма значение
- (“птиците” “д”) – допълнение, без друга информация
- (“зелените” “о” “мн”)
- (“жаби” “п” “мн”)
- (“гледат” “с” “мн”)
- (“към небето” “д”)

Смислено за нашата програма ще наричаме такова изречение, за което вторите етикети на подлога и неговото определение съвпадат, докато числата на подлога и сказуемите просто си „подхождат“ – ако подлогът е с етикет “**мн**”, значи и сказуемите трябва да е с “**мн**”; ако подлогът е с друг етикет за род, значи сказуемите трябва да е с етикет “**ед**”. Изреченията в нашия език ще могат да имат строго фиксирана структура, както показаните две – където всеки вид дума се среща точно по веднъж, и думите са подредени само в тази последователност. Освен това изреченията трябва да започват с главна буква и да завършват с точка – изисквания, които не зависят от участващите в него думи.

Целта на проекта е да се напише програма, която по зададена база от данни – списък от думи с техните коректни характеристики (етикети), генерира на произволен принцип смислени изречения, изградени от тях. Изреченията могат да бъдат само от вида, показан в двата примера по-горе, и трябва да спазва само нашите указания за „смисъл“ – да бъдат граматически коректни. Генерирането наведнъж на всички възможни изречения не е задължително изискване. Искане се всеки път, когато се извиква отново функцията (`getRandomSentence`) на конзолата да се изписва едно произволно такова.

Ако четенето на списъка с думи се извършва от файл, а генерираните изречения се записват в друг файл, то към оценката на проекта ще бъде прибавен бонус. Друга възможност за бонус е поддържането на втори вид изречения, които са по-дълги и имат структура като на “**Голямото куче и зелените жаби гледат към небето.**”. Тук критериите за смисъл на изречението са сказуемостта да е задължително в множествено число, а отделните двойки определение-подлог да съвпадат по същия начин. Родовете и числата на тези двойки нямат връзка едно с друго или със сказуемостта (както в дадения пример). В този случай отново искаме на произволен принцип да бъдат генерирани изречения и от двата вида.