

Оптимални обратими контекстно-свободни транслитерации

Димитър Скордев

Представяват известен интерес такива обратими контекстно-свободни транслитерации на български текстове, че при транслитерирането на българските букви да не се използват други символи освен латински букви. За простота нека се ограничим само с малките български и латински букви. Понеже първите са с четири повече от вторите, за да имаме обратимост на такава транслитерация, неизбежно е поне пет български букви да се транслитерират чрез буквосъчетания от две или повече латински букви. Разбира се, при това положение в общия случай дължините на текстовете ще се увеличават при транслитерирането. В някои случаи (например при изпращане на SMS) е желателно това увеличение да бъде възможно по-малко. Лесно е да се съобрази, че средната стойност на увеличението би била минимална тогава, когато петте най-рядко срещани български букви се транслитерират с двубуквени съчетания, а останалите български букви – с по една латинска буква.

Има основания да се приеме, че петте най-рядко срещани български букви са следните: **ш, ф, й, ю, ъ** (вж. Приложение 1 в [1]). Ако приемем това, то минимална средна стойност на увеличението на дължината при транслитерирането ще имаме за онези транслитерации от разглеждания вид, при които именно тези пет букви се транслитерират с двубуквени съчетания, а останалите 25 български букви – с по една латинска буква. Има много транслитерации от разглеждания вид, които удовлетворяват това условие, но измежду тях са за предпочитане онези, които са възможно по-естествени. Една не съвсем изкуствена такава транслитерация е например онази, която се определя чрез следната таблица:

а б в г д е ж з и й к л м н о п р с т у ф х ц ч ш щ ъ ь ю я
a b v g d e j z i i h k l m n o p r s t u p h f x c s h w y u h u h q

Като аргументация за направения при тази транслитерация избор в случая на българските букви **х** и **ц** би могло да послужи известно звуково сходство между техния прочит и прочита на съпоставените им латински букви съответно като „ф“ и „кс“. Довод за избора в случая на **ж** и **ч** може да бъде прочитът на съпоставените им латински букви съответно на френски и в определени случаи на италиански и румънски. Като оправдание за избора в случая на буквите **щ** и **я** може да послужи наличното леко графично сходство между тях и латинските букви, които са им съпоставени.

Литература

- [1] Скордев, Д., *Някои разглеждания във връзка с българските клавиатурни подредби*,
http://www.fmi.uni-sofia.bg/fmi/logic/skordev/bg_layouts.htm