Кириллица и латиница: игры со словами

Попалась мне на глаза довольно интересная книжка: Гик Е.Я. «Занимательные игры и развлечения». Прежде всего, там меня заинтересовал материал, касающийся применения компьютера для решения некоторых вопросов, связанных с играми со словами. Процитирую здесь некоторые своеобразные рекорды, установленные с использованием ЭВМ:

Самые длинные слова, являющиеся анаграммами друг друга (то есть состоящие из одних и тех же букв, но расположенных в разном порядке):

ТАРИФИКАЦИЯ – РАТИФИКАЦИЯ (11 букв);

ВЫБОРОЧНОСТЬ – ОБРЫВОЧНОСТЬ, МАГНИТОСФЕРА – СИНЕМАТОГРАФ (12 букв);

ПЕРЕМАЛЫВАНИЕ – ПЕРЕЛАМЫВАНИЕ (13 букв);

ОГРАНИЧЕННОСТЬ – НЕОРГАНИЧНОСТЬ (14 букв);

СТАРОРЕЖИМНОСТЬ – НЕРАСТОРЖИМОСТЬ, ВОДОСВЕТОБОЯЗНЬ – ОБЕЗЬЯНОВОДСТВО (15 букв);

ПЕРЕОРИЕНТИРОВКА – РЕПРОЕКТИРОВАНИЕ (16 букв).

Самые длинные слова, состоящие из разных букв: ЗВУКОСНИМАТЕЛЬ, РАЗГИЛЬДЯЙСТВО (14 букв); ЧЕТЫРЁХУГОЛЬНИК (15 букв).

Самые длинные слова, в которых гласные буквы чередуются с согласными: ВЕЛИКОМУЧЕНИЦА, СОЛОМОВОЛОКУША (14 букв) [1, с. 15-17].

В современных компьютерах для отображения букв (на экране или при выводе текста на печать) используются их изображения, хранящиеся в специальных файлах шрифтов. Для художественного оформления современными дизайнерами широко применяются шрифты типа TrueType (их файлы имеют расширение TTF), в которых изображение каждого символа (буквы) представляет собой векторный рисунок на основе кривых Безье второго порядка [2]. В операционных системах семейства Windows имеется определённый набор стандартных шрифтов, среди которых есть и относящиеся к вышеназванному типу. В частности это: "Times New Roman", "Arial", "Courier New", "Monotype Corsiva", "Script MT Bold" и "Vladimir Script".

Ни для кого не секрет, что некоторые буквы латинского алфавита по написанию совпадают с буквами алфавита русского. В связи с этим некоторые русские слова могут быть набраны при помощи латиницы – например, слово «орех» может быть составлено из латинских букв «о», «пэ», «е» и «икс» и на глаз определить такую подмену невозможно. Компьютер же здесь легко способен углядеть разницу, поскольку при работе с текстом он на машинном уровне имеет дело не с рисунками (изображениями букв), а с их двоичными кодами, которые для латинских букв строго и однозначно определены, а для национальных алфавитов есть довольно ограниченный набор кодов, обусловленных существованием разных кодовых таблиц. Для русского алфавита эти кодировки следующие: КОІ-8, Мас, ISO, DOS (кодовая страница 866) и Windows (кодовая страница 1251), кроме того, наши буквы имеются и в кодировке Unicode.

Существование разных кодировок для отображения текста на русском языке иногда создаёт проблемы. Каждый, кто регулярно пользуется Интернетом, наверняка хоть раз сталкивался с тем, что после загрузки веб-страница являет пользователю жуткую нечитабельную кашу из непонятных закорючек. Обычно это связано с тем, что браузер неверно определил кодировку, в которой нужно отображать содержимое загруженного html-файла (веб-страницы) и почти всегда эта проблема решается указанием вручную браузеру правильной кодировки (обычно это делается через соответствующую команду пункта меню «Вид»). Русским словам, набранным латинскими буквами, подобные неурядицы не страшны, однако несложно догадаться, что таких слов у нас не настолько много, чтобы из них можно было составить мало-мальски содержательный текст. Тем не менее, представляло интерес выяснить, какое наиболее длинное русское слово можно составить из латинских букв.

Указанная в начале заметки книга навеяла средство решения поставленной задачи с помощью простенькой программы, которая и была реализована на языке VBA в среде табличного процессора MS Excel. Фактически программа (макрос) производила поиск в заданном массиве слов, состоявших только из заранее оговорённого набора букв. Далее из сделанной компьютером выборки оставалось дополнительно отобрать самые многобуквенные экземпляры. В качестве исходного материала для обработки (поиска слов) была взята электронная версия «Полного орфографического словаря русского языка» под редакцией Лопатина В.В., скачанная когда-то с сайта "sheba.spb.ru".

Сначала был произведён поиск слов, состоящих из строчных букв «а», «е», «о», «р», «с», «у» и «х», поскольку они в подавляющем числе шрифтов (гарнитур) одинаковы по внешнему виду с латинскими буквами. Представленный набор довольно ограничен и поэтому ожидать очень длинных слов не приходилось. В результате программой было «выловлено» три слова из семи букв, являющихся в этом случае самыми длинными. Вот они:

accecop, paccoxa, peccopa

Слова можно писать и заглавными буквами. Буквы «А», «В», «Е», «К», «М», «Н», «О, «Р», «С», «Т» и «Х» совпадают на написанию с заглавными латинскими (строго говоря, во многих шрифтах русская «К» чуть-чуть отличается от латинской "К", но это различие принимается как несущественное и буквы считаются одинаковыми). По сравнению с первым случаем возможностей по словообразованию здесь больше, поэтому поиск по словарю даёт таких рекордсменов по длине:

```
АНАХОРЕТСТВО, ВЕРХОВЕНСТВО, ОРКЕСТРАНТКА, ОРКЕСТРОТЕКА, ОРНАМЕНТОВКА, ОСТЕОСАРКОМА, ОТВЕТСТВЕННО, СВЕРХОРКЕСТР, СВОЕВРЕМЕННО, СЕКРЕТАРСТВО, СКВАТТЕРСТВО, СОРОКАВАТТКА, СТАРОВЕРСТВО (12\ {\rm букb}); АВТОМОТОКРОСС, РАСТВОРОНАСОС (13\ {\rm букb}); СМЕНОВЕХОВСТВО, СООТВЕТСТВЕННО, ТЕРМОАНЕМОМЕТР (14\ {\rm буkb}); РЕСТАВРАТОРСТВО (15\ {\rm буkb}).
```

Если сосредоточиться на использовании только строчных букв, то здесь лучше подойдут менее распространённые гарнитуры. Так, шрифт "Monotype Corsiva" допускает написание "g", "m", "n", "u" взамен «д», «т», «п», «и» соответственно. Вот наиболее длинные слова, которые удалось отыскать в этом случае:

```
апостериори, археоптерис, партаппарат, пристрастие, стереотипер, супертрасса, троеперстие, туроператор (11 букв); периартериит (12 букв); радиооператор, радиорепортер (13 букв); аудиоаппаратура, радиоаппаратура (15 букв); стереоаппаратура (16 букв).
```

Некоторые шрифты имитируют рукописный текст, что даёт ещё больше дополнительных «степеней свободы»: в шрифте "Script MT Bold" буквы "b" и "r" изображаются подобно нашим «в» и «ч». В частности, получает возможность быть набранной на клавиатуре незабвенная чеховская «реникса» или просто

renyxa,

а самыми же многобуквенными здесь получаются следующие слова:

видеопиратство, простосердение, протопресвитер, псевдопериптер, радиодиспеттер, радиопиратство, регетворчество, староверчество (14 букв); видеоаппаратура, правотворчество, радиоаппаратура, реставраторство, стихотворчество, суперавтострада (15 букв);

стереоаппаратура (16 букв); страстотерпчество (17 букв).

В гарнитуре "Vladimir Script", вдобавок к остальным перечисленным буквам, буква "z" имеет такое же начертание, как русская «з». Компьютерный поиск наиболее длинных слов даёт такие результаты:

bugeoonepamop, bugeocpegembo, bozgyzonpobog, boenpouzbeemu, ocempobogembo, npeboezogembo, npegbozbeemue, npeebumepembo, neuxomepanebm, paguoonepamop, paguonepegara, paguonepexbam, paguopenopmep, penemumopembo, pempopazbumue, conpucumembue emuxombopembo uzupnamopembo ruemocepaerue

соприсутствие, стихотворство, узурпаторство, гистосердегие (13 букв);

bugeonupamembo, npoemocepgerue, npomonpecbumep, ncebgonepunmep, paguoguenemrep, paguonupamembo,

регетворгество, старовергество (14 букв);

bugeoannapamypa, boenpouzbogembo, npabombopreembo.

paguoannapamypa, peemabpamopembo, emuxombopreembo,

суперавтоетрада (15 букв);

nepenpouzbogembo, emepeoannapamypa, emepeoguanozumub,

еудопроизводетво (16 букв);

етраетотерптество (17 букв).

Одинаковость написания букв в двух разных алфавитах (русском и латинском) имеет и практическое применение в нашей стране: только заглавные буквы «А», «В», «Е», «К», «М», «Н», «О», «Р», «С», «Т», «Х», а также «У» (в силу определённой похожести на латинскую "Y") сейчас используются в автомобильных государственных регистрационных номерах. Это сделано для того, чтобы при выезде за границу на своём автомобиле где-нибудь в Германии или во Франции тамошние ДПС-ники (или как они у них там называются) при виде номера не вывихнули себе мозг от какойнибудь нашей «Ж» или «Щ».

Ссылки:

- [1]. Гик Е.Я. Занимательные игры и развлечения. М.: Дет. лит., 2010. 240 с.
- [2]. Статья «Кривая Безье» // RU.WIKIPEDIA.ORG: Википедия. Свободная энциклопедия. URL: https://ru.wikipedia.org/Кривая_Безье (дата обращения: 16.08.2020)

© Широков Александр, 16.08.2020