**Нахождение наиболее близкого значения**

Это глава из книги Билла Джелена [Гуру Excel расширяют горизонты: делайте невозможное с Microsoft Excel](http://baguzin.ru/wp/?p=12724).

Задача: люди вводят данные различными способами. Если вы попросите 50 торговых представителей записать прогноз для Дженерал Моторс, вы обнаружите дюжину различных способов написания названия компании или ее аббревиатуры. Объединить прогнозы от всех торговых представителей будет непросто. В столбце А (рис. 1) показаны некоторые способы написания имен клиентов, чьи официальные названия приведены в столбце D.

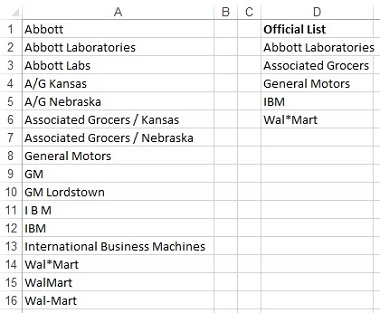


Рис. 1. Когда предложено ввести имена клиентов, сотрудники воспользуются разными написаниями или сокращениями

Это классическая задача известна как проблема нечетких соответствий. Она впервые обсуждалась на форуме MrExcel в 2001. В то время Хуан Пабло Гонсалес написал процедуру определения процента совпадений между двумя текстовыми строками. Повторно вопрос был поднят в 2003 г., когда Al\_B\_Cnu (ник) написал код функций FuzzyVLOOKUP, FuzzyHLOOKUP и FuzzyPercent. Поскольку код содержит около 400 строк, он здесь не приводится, но его можно найти в приложенном Excel-файле.

Решение: функция FuzzyPercent сравнивает текст из двух ячеек и определяет, какой процент символов *первой ячейки* присутствует в той же последовательности во второй ячейке (рис. 2). Здесь в ячейках А2 и В2 присутствует одинаковая последовательность из 11 символов. Поскольку ячейка A2 содержит 15 символов, то функция FuzzyPercent возвращает значение 73% (11/15=0,733). Обратите внимание, что, если поменять местами ячейки А2 и В2, то результат изменится. Это показано в строке 3. Здесь А3 и В3 также имеют общую последовательность из 11 символов, но теперь их часть в общем числе символов ячейки А3 снизилась до 50% (11/22=0,5).

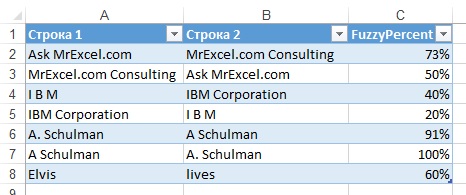


Рис. 2. Пользовательская функция FuzzyPercent вычисляет процент символов, которые находятся в той же последовательности

Обратите внимание, что алгоритм не совершенен. Строки *Elvis* и *lives* содержат в точности одни и те же символы, но в совершенно другом порядке. Действительно ли порядок совершенно другой? Нет. Обе строки включают символы *l-v-s* именно в такой последовательности. Т.е., 3 из 5 символов или 60% совпадают.

Проблема становится более сложной, когда вы должны найти наилучшее соответствие в двух списках (рис. 3). Здесь двумерная таблица показывает, насколько хорошо каждое название из реального списка (столбец А) совпадает с официальным названием (строка 1). Отметим, что *GM Lordstown* имеет одинаковое число совпадений, как с *General Motors*, так и с *Abbott* *Laboratories*. Это не очень удачно, поэтому был предложен альтернативный алгоритм, реализованный в функции FuzzyPercent2().[[1]](#footnote-1)

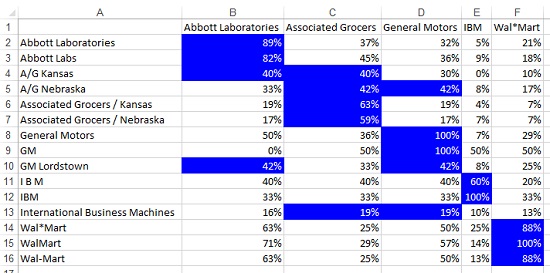


Рис. 3. Наилучшее соответствие для каждой строки выделено цветом

Al\_B\_Cnu также написал функции аналогичные ВПР и ГПР, но для поиска наилучшего неполного соответствия. Как и ВПР, FuzzyVLOOKUP возвращает определенный столбец из таблицы подстановки. Также она может выдать первое наилучшее соответствие, второе и так далее (рис. 4).

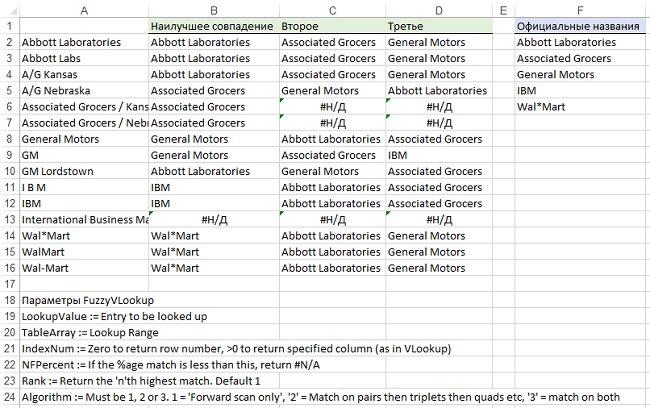


Рис. 4. FuzzyVLOOKUP сравнивает каждое название в столбце А со списком официальных названий (столбец F), и возвращает наилучшее соответствие, второе лучшее, третье лучшее

Резюме: нахождение наиболее близкого значения может быть выполнено с помощью пользовательской функции VBA.

[Источник](http://www.mrexcel.com/forum/excel-questions/69491-closest-match.html).

1. У меня функция FuzzyPercent2() не заработала. – *Прим. Багузина* [↑](#footnote-ref-1)