

Автор: Виктор (Jyj, Жы Жъ) сайт автора <http://www.liveinternet.ru/users/jyj/>, форум «Руны Русского Рода» <http://prarod.forum2x2.ru/>

Анализ двоичного слова

Предисловия я писать не буду, так как по этой теме существует достаточно материала. В этом материале есть и предисловие и обоснование того зачем необходим анализ слова. Поэтому будем говорить по существу, то есть, я изложу свой вариант анализа слова.

Слова состоят как минимум из двух групп букв. Одна группа – согласные. Другая – не согласные (Не согласные и несогласные, вот воистину красота Русского языка. Один только знак – пробел, и смысл, и суть, всё изменилось). В русском алфавите 10 букв не согласных и 23 буквы, согласных. Другого не дано. Этим и воспользуемся. Пусть, все гласные (не согласные) буквы объединённые в одну группу, будем обозначать их буквой Г, а согласные – буквой С. Тогда слово «привет» можно записать следующим образом: ССГСГС. Любое русское слово можно записать в таком виде.

Понятно, что смысл слова «привет» при переходе к записи «ССГСГС», потерял, но похоже, что-то осталось. Вот с этим что-то, мы и будем разбираться на начальном этапе.

Представим, что некий паучок, соткал паутину и движется только в плоскости этой паутины. Ячейки этой паутины - правильный квадрат. Двигаться паучок может только вдоль паутинок, не иначе. Однажды паучок решил перебраться из красной точки в жёлтую точку (Рис.1).

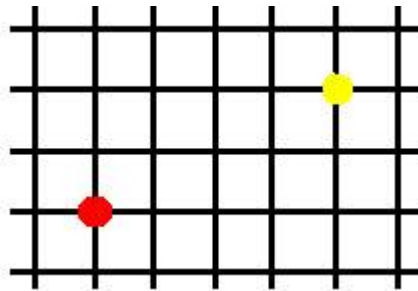


Рис . 1.

Паучок, будучи сметливым, чему научила его жизнь, выбирает только одно движение – вперёд к цели. Ни одного лишнего шага. То есть, по горизонтали только вправо, по вертикали только вверх. В данном случае, когда по горизонтали 4 ячейки и по вертикали 2, паучок может пойти, в том числе по следующему маршруту – синяя линия идущая от красной точки к жёлтой (Рис. 2).

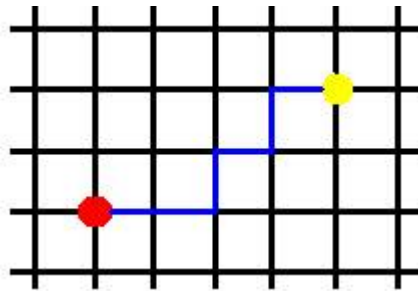


Рис. 2.

Пусть, движение паучка по горизонтали обозначается буквой - «с», а по вертикали буквой «г». Тогда его маршрут запишется очень просто «ссгсгс». Надеюсь, вы поняли к чему я пришёл? Я пришёл к абсолютной модели, которая описывает русские слова, естественно их нужно сначала перевести в виде гласных-согласных. Вопреки устоявшемуся понятию, что такое двоичное представление, я такое представление русских слов также назову двоичным, только числа, которые для этого понадобятся будут 1 и 2. И вот почему. Слово «привет» - «ССГСГС» можно представить в виде «112121», а это последний шаг к переводу его в число

$$\text{привет} \rightarrow \text{ССГСГС} \rightarrow 112121 \rightarrow 1 * 3^5 + 1 * 3^4 + 2 * 3^3 + 1 * 3^2 + 2 * 3 + 1 = 394$$

Так как старший разряд всегда значим, и он не может быть равным нулю - потеряется первая буква, то использование «0» исключено. Наши русские слова начинаются как с гласных, так и с согласных. Паучок может первый шаг сделать как по горизонтали, так и по вертикали. Следовательно, следующим по порядку за нулём только 1 и 2, а система исчисления троичная – принцип минимизации, как у паучка. Выше приведённый многочлен имеет в десятичной системе исчисления конкретное число равное 394. Ставится вопрос: А что 1 и что 2? Опять же сметливость паучка или принцип наименьшего действия, т.е. всё что угодно, но по минимуму количества, и не в ущерб (максимум) качества. Статистические данные показывают, что в русских словах гласных меньше чем согласных, и если мы словам будем приписывать характеризующие их числа, то желательно чтобы эти числа были как можно меньше, поэтому выбираем: Гласные буквы соответствуют 2. Согласные буквы соответствуют 1. В противном случае наш пример будет выглядеть следующим образом:

$$\text{привет} \rightarrow \text{ССГСГС} \rightarrow 221212 \rightarrow 2 * 3^5 + 2 * 3^4 + 1 * 3^3 + 2 * 3^2 + 1 * 3 + 2 = 698$$

Наш паучок, гипотетически, лазить может и по катетам, тогда мы получим, ну например ссссгг или в двоичном представлении слова - 222211. В русском языке таких слов нет, разве что чиновничья мысль может породить столь содержательную аббревиатуру. И наш паучок, руководствуясь

генеральной линией сей аббревиатуры, сначала побредёт вдоль одного катета, поглядывая с воодушевлением на цель, потом вдоль другого катета, если не пропустит момента, когда нужно свернуть на 90 градусов, то благополучно доберётся до цели. Осмелюсь заверить вас, паучок ну никак по катетам двигаться к своей цели не будет. Что же двигает паучком? Ответ: ощущение цели! Паучок со своим дискретным пространством свыкся, породнился, прикипел. Он реального своего пространства не замечает. Для него есть цель, направление на цель, следовательно, линейность во всех направлениях, и в его паучьих мозгах тоже линейность, хотя реальность иная. Бедняга и не догадывается, что не зависимо от того каким путём он достигается цели, расстояние во всех случаях одинаковое. Главное процесс! Но каков процесс! Там, повернул, там пробежался, опять повернул. Так, что чиновники тоже в чем-то правы, на то они и чиновники, чтобы правИТЬ право. Т.е. как бы цель для всех одна и для народа и для чиновников, только народ желает идти к цели напрямик, а чиновник посылает его по катетам. Мол, не засоряйте себе голову - идите прямо, а когда повернуть налево, или направо мы вам скажем. Народ пришёл к цели, получил полный комплект букв, который составляет Русский алфавит, в том числе букву – ё. Ему говорят:- «Нет! Ё лишнее, и не надо её защищать, сажайте лен». Вот мы вместо того, чтобы сажать и возвращать лён, Лен и насажали.

Примечание:

Коэффициент отношения дискретного пространства к линейному равен:

$$k = \sqrt{1 + \frac{2ab}{a^2 + b^2}}$$

где a и b – длины катетов. Готов обсуждать это примечание, но не в рамках этой работы.

И так, всё вроде логично, один момент – поведение паучка. Не буду раскрывать эту тему в данной работе, её подробно и красиво раскрыл Ричард Фейнман («Фейнмановские лекции по физике» Том 6. Глава 19: Принцип наименьшего действия). Правда, он говорил не о паучке. Он говорил о Природе, а так как она Всё, то и паук тоже. Ещё хотел обратить внимание читателя на «Принцип наименьшего времени Ферма». В 1650 году этот принцип Пьер Ферма использовал для наглядного объяснения закона поведения света. Свет выбирает из всех возможных путей, соединяющий две точки, тот путь, который требует наименьшего времени для его прохождения. Этот принцип и закон в нашем случае относятся к ощущениям паучка. В конце концов, есть же у паучка глаза...

Далее сразу же переходим к интегралам. Кто не знает, не волнуйтесь, в конце будут простые формулы. Куда же без них, на то он и анализ, что математический, что иной.

Выдержка из книги Гильберта Стренга «Линейная алгебра и её применение» (Издательство «Мир», Москва 1980 г. § 3.3. Ортогонализация Грама – Шмидта, стр. 163).

Пусть на отрезке $0 \leq x \leq 1$ задана функция и мы хотим приблизить её прямой линией $y = C + Dx$

Существуют по крайней мере три способа (я изложу из этой книги один способ) отыскания этой наилучшей прямой....

(i) Минимизировать

$$E^2 = \int_0^1 (\sqrt{x} - C - Dx)^2 dx = \frac{1}{2} - \frac{4}{3}C - \frac{4}{5}D + C^2 + CD + \frac{D^2}{3}$$

Приравнивая нулю производные, получаем

$$\frac{\partial E^2}{\partial C} = -\frac{4}{3} + 2C + D = 0$$

$$\frac{\partial E^2}{\partial D} = -\frac{4}{5} + C + \frac{2D}{3} = 0$$

$$C = \frac{4}{15}, D = \frac{4}{5}$$

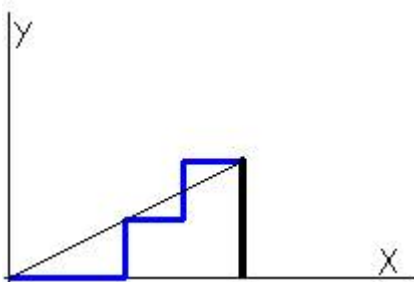
Отсюда получим решение

Искомый вид уравнения будет

$$y = \frac{4}{15} + \frac{4}{5}x$$

О чём это говорит. На участке от 0 до 1, сумма квадратов всех отклонений искомой функции от заданной минимальна. Следовательно, найденная функция (из заданного класса функций) максимально полно описывает заданную функцию.

А что у нас? У нас следующее.



Нам всё задано.

По оси откладывается количество согласных в слове. По оси у количество гласных в том же слове, получаем уравнение гипотенузы

$$y = \frac{\Gamma}{C}x$$

Где Γ – количество гласных, C – количество согласных букв в анализируемом двоичном слове. В слове «привет» или сгсгс: $\Gamma=2$, $C=4$. Гипотенуза, идеальный путь для нашего паучка, он старается его придерживаться. Уравнение гипотенузы имеет вид

$$y = \frac{2}{4}x \rightarrow y = \frac{1}{2}x$$

$$0 \leq x \leq C = 4$$

Это первое уравнение.

Следующее уравнение не одно. Количество уравнений столько сколько согласных. Уравнения всех горизонтальных синих линий

$$y_1 = 0 \quad 0 \leq x \leq 1$$

$$y_2 = 0 \quad 1 \leq x \leq 2$$

$$y_3 = 1 \quad 2 \leq x \leq 3$$

$$y_4 = 2 \quad 3 \leq x \leq 4$$

Вертикальные синие линии, как ни странно не учитываются, хотя именно они уведат нашего паучка от истинной линии. Объяснять, почему не учитываются - не буду, так как это уже чистая математика, а не она в этой работе главное. Математика – инструмент, причём надёжный.

Запишем формулу квадрата полного отклонения для заданных функций

$$E^2 = \int_0^1 \left(0 - \frac{1}{2}x\right)^2 dx + \int_1^2 \left(0 - \frac{1}{2}x\right)^2 dx + \int_2^3 \left(1 - \frac{1}{2}x\right)^2 dx + \int_3^4 \left(2 - \frac{1}{2}x\right)^2 dx$$

$$E^2 = \frac{5}{6}$$

Паучок, проползая от красной точки к жёлтой, с целью передать слово «привет», допустим своей подружке, отклонялся от идеального пути. Суммарное отклонение составило $5/6$.

Обобщим, т.е. выведем формулу для любого двоичного слова

$$E^2 = \sum_{i=1}^C \int_{i-1}^i \left(g[i] - \frac{\Gamma}{C}x\right)^2 dx = \frac{C\Gamma^2}{3} + \sum_{i=1}^C \left\{g[i]^2 - 2\frac{g[i]\Gamma}{C} + \frac{g[i]\Gamma}{C}\right\}$$

Формула 1

Где Γ – количество гласных в двоичном слове, C – количество согласных в двоичном слове, $g[i]$ – целые положительные числа, количество которых равно C . Для этих чисел справедливо неравенство

$$0 \leq g[i] \leq C$$

Учитывая сметливость нашего паучка, а именно, он движется по вертикали только вверх, то справедливо неравенство

$$0 \leq g[1] \leq g[2] \leq \dots \leq g[C] \leq C$$

Неравенство 1

Формулу 1 запишем в другом виде

$$E^2 = D + P$$

Где

$$D = \frac{C^2}{3} + \sum_{i=1}^C \left\{ g[i]^2 + \frac{g[i]i}{C} \right\}$$

$$P = -2 \frac{\Gamma}{C} \sum_{i=1}^C g[i] * i$$

Значение D – как вы уже успели заметить, абсолютно не зависит от того в каком порядке идёт суммирование заданных величин. Что $5+4=9$, что $4+5=9$ без разницы, а вот значение P , зависит от того в каком порядке происходило суммирование, так как под знаком суммы есть два множителя $g[i+1]$ и i .

Есть разница между вариантами?

$$5*1+4*2=13$$

$$4*1+5*2=14$$

Есть. Множители 1,2,3 и т.д. учитывают порядок включения в общую сумму чисел - 4, 5 и т.д.

В нашем примере для паучка $E^2 = \frac{8}{6}D$ и P будут следующие

$$D = \frac{4 * 2^2}{3} + \left(0^2 + \frac{0 * 2}{4} \right) + \left(0^2 + \frac{0 * 2}{4} \right) + \left(1^2 + \frac{1 * 2}{4} \right) + \left(2^2 + \frac{2 * 2}{4} \right) = \frac{71}{6}$$

$$P = -2 \frac{2}{4} (0 * 1 + 0 * 2 + 1 * 3 + 2 * 4) = -11$$

Что вместе составляет $\frac{71}{6} - 11 = \frac{8}{6}$

Поможем паучку найти самый оптимальный вариант передвижения от красной точки к жёлтой, т.е. найдём минимальное значение величины E^2 . В формуле 1, Γ и C постоянные, изменяться только $g[l]$ по ним и про дифференцируем

$$\frac{\partial E^2}{\partial g[l]} = 2g[l] - 2\frac{\Gamma}{C}l + \frac{\Gamma}{C} = 0$$

$$g[l] = \frac{\Gamma(2l - 1)}{2C}$$

Формула 2

где l — натуральное число, изменяющееся в диапазоне $1 \leq l \leq C$

Итак, при условии, что $\Gamma=2$, $C=4$, получаем

$$g[1] = \frac{2(2 * 1 - 1)}{2 * 4} = \frac{1}{4}$$

$$g[2] = \frac{2(2 * 2 - 1)}{2 * 4} = \frac{3}{4}$$

$$g[3] = \frac{2(2 * 3 - 1)}{2 * 4} = \frac{5}{4}$$

$$g[4] = \frac{2(2 * 4 - 1)}{2 * 4} = \frac{7}{4}$$

Естественно, наш паучок так двигаться не может, ячейка его паутинки квадрат 1×1 . Однако не зависимо от этого мы можем представить, что подружка паучка, сплела иную паутину, ячейка которой прямоугольник $1 \times 1/4$ (по горизонтали 1, по вертикали $1/4$). Вот она то и сможет пробежать по самому оптимальному варианту к своему паучку. Структура передвижения паучихи - гсгсгсгсгсгсг, где один шаг по вертикали равен $1/4$. По вертикали она опускается только вниз. По горизонтали только влево. Если под её структурой передвижения понимается двоичное слово и далее слово, то такая структура напоминает слова из языковых групп семитов и угоро-финов. Много гласных.

Примечание:

Наиболее очевидное различие между гласными и согласными с артикуляционной точки зрения. При образовании гласных звуков речевой канал, и воздух, выходящий из легких, почти не встречает преград на своем пути. Поэтому для гласных характерна слабая струя воздуха. При этом степень раствора может быть различной (большая при произнесении А, меньшая при произнесении Э и О). При образовании согласных струя воздуха, выходящая из легких, встречает в гортани или в надставной трубе препятствия (сужение речевого канала или затвор его). Согласные возникают при преодолении различных преград. В связи с необходимостью преодолевать преграды при образовании согласных наблюдается более сильная воздушная струя. При образовании гласных мускульное напряжение почти в равной мере распространяется по всему речевому аппарату. Артикуляция не локализована. При образовании согласных преодоление преград вызывает сосредоточение мускульного напряжения в том месте, в котором преграда. Следовательно, для согласных характерна

локализованная артикуляция, т.е. все напряжение органов произношения сосредотачивается в одном месте. Гласные же характеризует высота резонаторного тона (самый высокий звук И, а самый низкий У), а также артикуляцией, объемом (регулируется движением губ) и формой резонатора. Гласные в русском языке являются основой слога, и могут быть ударными или неударными. При произношении согласных, организм затрачивает больше энергии, чем при произношении гласных. Язык, в словах которого много гласных находится ещё на стадии развития, т.е. освоения согласных. Для освоения согласных необходимо затрачивать больше энергии.

Паутины паучков при наложении друг на друга в некоторых узлах совпадают, только паучок чувствует свои нити и свои узлы, а паучиха свои нити, узлы нитей паучка принимает за свои. А У СВЕТА СВОЙ РАСКЛАД! Это я опять возвращаю вас к ощущениям паучка.

Тогда минимальное отклонение, которое возможно, при перемещении из красной точки в жёлтую точку, по дискретной сетке (при условии, что катеты равны 4 и 2),

имеет значение

$$E^2 = \frac{2^2 * 4}{3} + \left(\left(\frac{1}{4} \right)^2 - \frac{2 * \frac{1}{4} * 2 * 1}{4} + \frac{\frac{1}{4} * 2}{4} \right) + \left(\left(\frac{3}{4} \right)^2 - \frac{2 * \frac{3}{4} * 2 * 2}{4} + \frac{\frac{3}{4} * 2}{4} \right) + \left(\left(\frac{5}{4} \right)^2 - \frac{2 * \frac{5}{4} * 2 * 3}{4} + \frac{\frac{5}{4} * 2}{4} \right) + \left(\left(\frac{7}{4} \right)^2 - \frac{2 * \frac{7}{4} * 2 * 4}{4} + \frac{\frac{7}{4} * 2}{4} \right) = \frac{1}{12}$$

$$D = \frac{151}{12}$$

$$P = -\frac{25}{2}$$

Примечание:

Если Вас, чиновник, послал по катетам 4 на 2, т.е. четыре двери прямо по коридору и налево вторая дверь, то $E^2 = \frac{16}{3}$ $D = \frac{16}{3}$ $P = 0$

Самое максимальное отклонение, но зато идеальный порядок.

Проверку проведите сами для того, чтобы убедиться в том, что мой метод расчёта усвоили хорошо.

Но как же быть паучку со своими квадратами 1x1? Условимся, что если g[i] не целое число, то будем его округлять следующим образом: Если мантисса числа g[i] больше 0.5 то берём число целое больше чем g[i], если меньше 0.5 то берём число целое меньше g[i]. Для нашего паучка получим

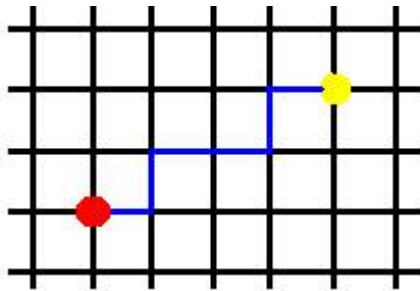
$$g[1] = \frac{1}{4} \rightarrow g[1] = 0$$

$$g[2] = \frac{2(2 * 2 - 1)}{2 * 4} = \frac{3}{4}$$

$$g[3] = \frac{5}{4} \rightarrow g[3] = 1$$

$$g[4] = \frac{7}{4} \rightarrow g[4] = 2$$

Самый оптимальный путь для нашего паучка следующий - сгсгсг



$$Eo^2 = \frac{2^2 * 4}{3} + \left((0)^2 - \frac{2 * 0 * 2 * 1}{4} + \frac{0 * 2}{4} \right) + \left((1)^2 - \frac{2 * 1 * 2 * 2}{4} + \frac{1 * 2}{4} \right) + \left((1)^2 - \frac{2 * 1 * 2 * 3}{4} + \frac{1 * 2}{4} \right) + \left((2)^2 - \frac{2 * 2 * 2 * 4}{4} + \frac{2 * 2}{4} \right) = \frac{1}{3}$$

$$Do = \frac{40}{3}$$

$$Po = -13$$

Оптимальное двоичное слово которое паучок может передать своей подружке – сгсгсг или например «космос».

Вычислив оптимальный путь для паучка, отложим его в сторону и выведем общую формулу для оптимального $E^2 = dE$ при перемещении в дискретном пространстве. Для этого в формулу 1 подставим значение $g[i]$ из формулы 2

$$dE = \frac{r^2}{12C}$$

$$dD = \frac{r^2(8C^2 - 1 + 6C)}{12C}$$

$$dP = -\frac{r^2(C+1)(4C-1)}{6C}$$

Представим последние три формулы в виде чисел, для наглядности. Таблицы 33x33 смотрите ([Таблица dE](#) [Таблица dD](#) [Таблица dP](#)).

Я надеюсь, что вы обратили внимание на то, что рассчитанные оптимальные структуры при передвижении из красной точки в жёлтую точку для паука – сгссгс и паучихи (в дискретном пространстве $1 \times \frac{1}{4}$) – гсгсгсгсгсгсг обладают центральной симметрией. Для того, чтобы как-то количественно оценить качество структуры исследуемого двоичного слова необходимо его с чем-то сравнить. Самое разумное, сравнить с некой упорядочной структурой. В нашем случае, для двоичных слов имеющий одинаковый набор гласных и согласных: ссгсгс, сгссгс и т.д. оптимальной (наименьшая ошибка) упорядочной структурой будет – сгссгс. Сравнение будем осуществлять в виде отношения, которое назовём **коэффициентом линейности**

$$k = \frac{E\sigma^2}{E^2}$$

Где $0 < k \leq 1$. Чем ближе k к 1 тем линейнее двоичное слово. Для слова «привет», коэффициент линейности будет равен

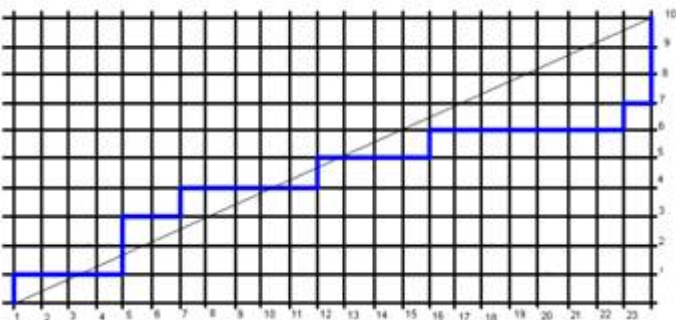
$$k = \frac{\frac{1}{3}}{\frac{1}{5}} = \frac{2}{5} = 0.4$$

В общем виде формула для коэффициента линейности будет иметь вид

$$k = \frac{\Gamma^2 C^2 - 3 \sum_{i=1}^C g g[i] (2\Gamma i - g g[i] C - \Gamma)}{\Gamma^2 C^2 - 3 \sum_{i=1}^C g [i] (2\Gamma i - g [i] C - \Gamma)}$$

Где $g[i]$ – значения исследуемого двоичного слова, $gg[i]$ – значения оптимального двоичного слова, рассчитанные по формуле 2 с условиями округления. Слова «привет» и «космос» имеющие структуры ссгсгс - $g=[0,0,1,2]$ и сгссгс - $gg=[0,1,1,2]$, соответственно, соотносятся друг к другу как дробь 2/5.

На завершающем этапе этой работы рассмотрим ещё один пример. Наш Русский алфавит, в том виде как нас научили его произносить ещё в школе, можно представить в виде слова – абвгдеёжзийклмнопрстуфхцчшщъызьёя. Переведём его в двоичное слово – гсссгсгсгсгсссгссссгссссгсггг. Графически это будет выглядеть следующим образом



Задано $g[i]=[1,1,1,1,3,3,4,4,4,4,4,5,5,5,5,6,6,6,6,6,6,7]$, $\Gamma=10$, $C=23$.

Используя формулу 2, получим $gg[i]= [0,1,1,2,2,2,3,3,4,4,5,5,5,6,6,7,7,8,8,8,9,9,10]$

формуле k_0 для оптимального слова. Коэффициент линейности для дискретного пространства двоичных слов будет равняться

$$k = \frac{k_0}{k_1}$$

Далее, структура двоичного слова может обладать внутренней симметрией относительно центра слова (сггсс) , а также упорядочной повторяемостью (гсгсс).