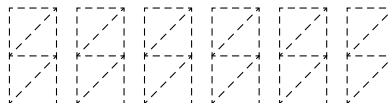
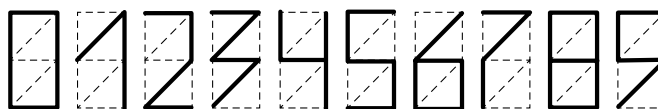


0123456789 Идеальный ПОЧТОВЫЙ ИНДЕКС

В далёком 1971 году Министерство связи (тогда ещё СССР) ввело в обращение почтовые индексы. Иными словами, каждому почтовому отделению был присвоен свой уникальный шестизначный числовой индекс, а на конвертах появился специальный шаблон – так называемый *кодовый штамп*, содержащий заготовки для шести цифр:



Надо было обвести синим или чёрным цветом нужные линии, чтобы образовались цифры. Но обводить не как кому нравится, а в соответствии с образцом, который имелся на обратной стороне конверта:



При обработке корреспонденции конверты пропускались через сортировальный автомат с оптическим датчиком, который и определял их дальнейшую судьбу – куда какой конверт направится. Такое нововведение позволило существенно ускорить доставку писем (доказательством эффективности системы является тот факт, что она по сей день используется, например, в России).

Но *почему* изображения цифр имеют именно такой вид, как на образце? Ведь большинство из них можно было нарисовать и по-другому. Неужели они создавались неведомым нам дизайнером, что называется, «от фонаря», лишь бы внешне напоминали привычные символы?

Одним из тех, кто попытался дать научно обоснованный ответ на этот вопрос, был Ярослав Карпов (тогда ещё десятиклассник из тогда ещё Ленинграда), опубликовавший на страницах 11-го номера «Кванта» за 1987 год статью «Оптимальная кодировка почтового индекса». Исходил он из следующих предположений. Сортировальный автомат может, хотя и с очень малой вероятностью, ошибаться, причём не только и даже не

столько из-за неисправности самой техники, сколько из-за небрежности человека при заполнении кодового штампа (искривлённые линии, слишком бледный цвет и т.д.). Если автомат ошибётся при распознавании какой-либо цифры, то это может привести к одному из двух исходов:

1) Воспринятое устройством ошибочное изображение не совпадет ни с какой из других цифр. Это, конечно, неприятно, но не фатально – такое письмо будет перенаправлено на ручную сортировку, и задержка при прохождении корреспонденции окажется не слишком большой.

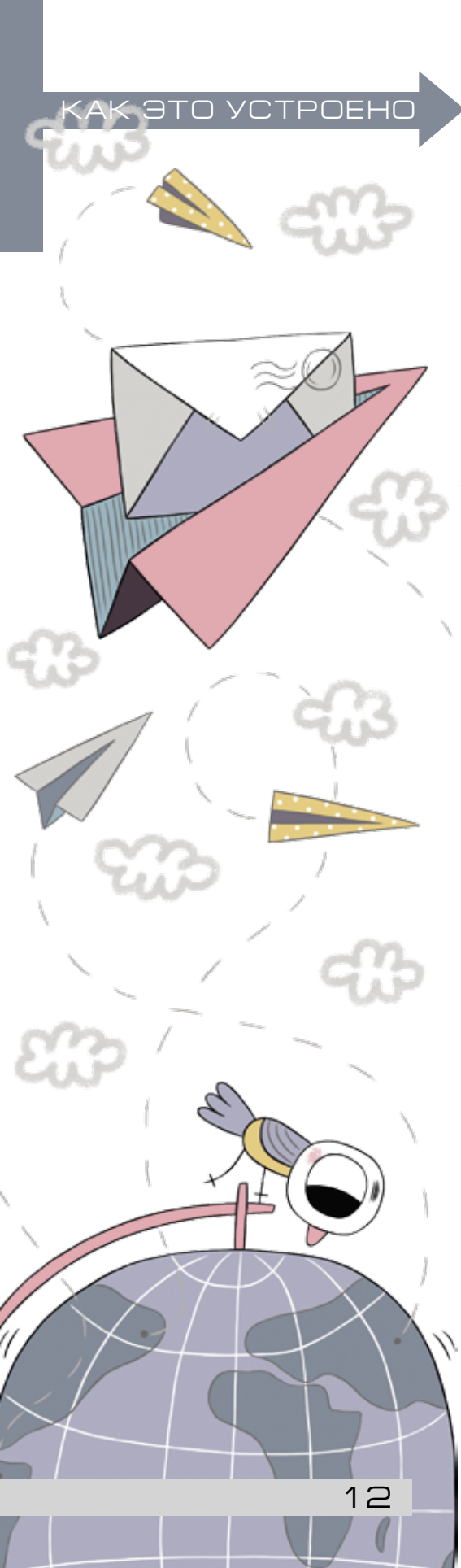
2) Воспринятое устройством ошибочное изображение *совпадёт* с какой-то из других цифр. Это гораздо хуже, потому что письмо будет отправлено в другое место. Учитывая географические масштабы государства (тогдашнего, да и нынешнего), нетрудно понять, что пока разберутся и всё поправят – пройдёт немало времени.

Поэтому важнейшей задачей разработчиков системы индексов должно было быть сведение к минимуму вероятности принять одну цифру за другую. Возможно, с этим и связан именно такой внешний вид цифр в индексах?

Чтобы проверить свою догадку, Я. Карпов сначала изобразил всевозможные способы приемлемого изображения цифр. Вот что у него получилось:

Цифра 0		Цифра 5	
Цифра 1		Цифра 6	
Цифра 2		Цифра 7	
Цифра 3		Цифра 8	
Цифра 4		Цифра 9	





Из них можно составить всего $1 \cdot 2 \cdot 4 \cdot 3 \cdot 2 \cdot 2 \cdot 2 \cdot 3 \cdot 1 \cdot 2 = 1152$ варианта десяти цифровых «комплектов». Далее Карпов ввёл понятие «расстояние между кодировками», равное количеству несовпадающих отрезков при изображении различных цифр. Возьмём, например, принятое в настоящее время изображение цифр 2 и 5. Расстояние между кодировками этих цифр равно 5, так как у них имеется 5 несовпадающих отрезков (т.е. таких, что у одной из цифр соответствующий отрезок проведён, а у другой – нет). А, скажем, расстояние между кодировками цифр 0 и 8 равно лишь 1, ибо у них расхождение имеет место в единственном отрезке – среднем горизонтальном.

Обозначим через p вероятность принятия автоматом проведённого отрезка за «непроведённый» или наоборот. Тогда вероятность правильного распознавания одного отрезка равна $1 - p$. Вероятность же принять одну цифру за другую равна $p^k \cdot (1 - p)^{9-k}$, где k – расстояние между кодировками цифр, а 9 – общее количество отрезков. Оно и понятно – чтобы спутать одну цифру с другой, необходимо, чтобы ровно k отрезков (как раз те, в которых изображения различаются) были восприняты с ошибкой, а остальные $9 - k$ отрезков автомат прочитал верно. Таким образом, устройство принимает двойку за пятёрку (или же пятёрку за двойку) с вероятностью $p^5 \cdot (1 - p)^4$. Ну, а восьмёрка и ноль имеют шансы быть перепутанными с вероятностью $p \cdot (1 - p)^8$.

Что же далее? Осталось задаться каким-либо «разумным» значением p , после чего для каждого из 1152 наборов найти вероятность спутать каждую пару цифр (а таких пар, как легко видеть, $10 \cdot 9 / 2 = 45$) и просуммировать все эти вероятности. Полученную сумму можно считать критерием «качества»: чем она больше, тем меньше вероятность отправления письма в другое место при использовании того или иного набора.

Я. Карпов всё это проделал, используя компьютер (тогда ещё ЭВМ), для различных значений p . Оказалось, что при $p \leq 0,3$ наилучшим является именно тот набор, который используется в настоящее время.

А поскольку реальное значение p заведомо не превышает 0,3 (иначе грош цена такой технике), то получается, что разработчики системы шли по такому же пути!

«Конечно, мне было бы приятнее, – отмечает напоследок Я. Карпов, – если бы моя программа выявила не принятый на почте набор, а другой: я тогда смог бы предложить заменить принятый набор на свой – лучший». К сожалению, не судьба!

...Однако с тех пор многое изменилось. Прежде всего, появилось множество световых табло в самых разных местах (лифты, светофоры, часы). И как-то автору этой заметки попало на глаза другое изображение цифры 1, существенно отличающееся от тех, что «обсчитывал» тогдашний десятиклассник. Вот оно:



Как видно, здесь используются два левых вертикальных отрезка, а не правых. И это изображение ничуть не хуже двух других. А раз так, то имеем возможность получить дополнительно к рассмотренным еще 576 наборов. А дальше – дело техники (в данном случае вычислительной). Проверка на компьютере по методике Карпова показала, что существует *два* набора, которые превосходят используемый в настоящее время. В лучшем из них все цифры изображаются так же, как и ранее, кроме этой самой единицы. Во втором, который чуть хуже (но всё-таки превосходит нынешний), кроме единицы по-другому изображается четвёрка (см. изображения выше). Так что Ярослав, к великому сожалению, чуть-чуть не дотянул до того, чтобы превзойти используемый набор. Обидно!

И хотя справедливость в конце концов восторжествовала (улучшение всё-таки оказалось возможным), обращаться в Министерство мы не станем. Во-первых, переучить миллионы людей – это ох как непросто, а во-вторых – зачем? Ведь, как ни крути, обычная почта в конвертах уверенно и неизбежно уступает место электронной переписке. Ибо прогресс не остановить. //



Художник Наталья Гаврилова

