

Размышления. Количество информации

Один мудрец взял пустой кувшин и наполнил его доверху небольшими камнями. Собрал своих учеников и задал им первый вопрос: "Скажите, уважаемые, полон ли мой кувшин" На что те ответили: "Да, полон". Тогда мудрец взял полную банку с горохом и высыпал содержимое в кувшин с камнями. Горох занял свободное место между камнями. Задал мудрец второй вопрос: "Полон ли теперь мой кувшин" Ученики вновь подтвердили, что полон. Тут мудрец взял коробку с песком и его тоже высыпал в кувшин. Песок просочился сквозь горох и камни и занял все свободное место и все закрыл. Еще раз спросил мудрец своих учеников, полон ли кувшин и снова услышал утвердительный ответ. Тогда достал мудрый человек кружку, полную воды и вылил ее в кувшин до последней капли.

Физический смысл понятия "энтропия" часто вызывает затруднение. Вероятно, наилучшей аналогией будет "плотность" или, может быть, "структурная плотность". Предельно плотная упаковка, "100% чистого вещества" отвечает энтропийному пределу: невозможно очистить продукт от примесей лучше, и чем ниже содержание искомого продукта в смеси, тем ближе к нулю его энтропия.

Нелепо приписывать оценке Хартли-Шеннона какую-либо связь с "информативностью" сообщения - в обычном, обыденном смысле этого слова. Исторически сложившееся название "количество информации" следует признать скорее маркетинговым слоганом, нежели сколько-нибудь реальной характеристикой.

Иронично, но вычисленное по формулам "количество информации", к информации ни малейшего отношения не имеет.

Если рассмотреть, например, процедуру взвешивания на рычажных весах, то килограмм сахарного песка может быть уравновешен килограммом золотого песка. Ни по каким своим характеристикам два этих товара не равноценны друг другу: золотой песок бессмысленно сыпать в чай, а сахарный - нести к ювелиру.

Но если требуется переслать и тот и другой посылкой с оплатой по весу, то в этом, и только в этом смысле они оказываются эквивалентны.

Оценки Хартли-Шеннона - это оценки платы за транспортировку сообщения, безотносительно к тому какая в нем содержится информация и содержится ли вообще: транспортировка дезинформации или шума, например, обойдется в ту же самую сумму.

Энтропия сообщения - это технический показатель его "насыпной плотности", "чистого веса", "нетто", позволяющий рассчитать предельное, минимально необходимое для транспортировки количество тары заданной вместимости, и не учитывающий ни природу товара, ни его ценность для получателя.

Очевидно, для расчета потребной тары (пропускной способности канала связи) существенна только несжимаемая часть - энтропия, которая и будет выполнять роль

меры.

Важно понимать, что выражения вида "энтропия - мера неопределенности" или "энтропия - мера беспорядка" не более, чем не имеющие смысла "маркетинговые уловки". И, тем более, энтропия никак не связана с информационным содержанием сообщения.

Аксиоматически, будучи именно мерой, энтропия совершенно безразлична к порядку и соотношению своих составляющих, важна только их сумма. Шенноновская энтропия сообщения $0, 1, 2 \dots N$ не изменится от любой, сколь угодно случайной перестановки ("перемешивания") символов последовательности. По критерию энтропии неразличимы любые анаграммы. Например, в парах

ля - ял

хан - хна

сорт - трос

карта - карат

приказ - каприз

мошкара - ромашка

апельсин - спаниель

лекарство - стекловар

дозревание - раздвоение

ратификация - тарификация

австралопитек - ватерполистка

нерасторжимость - старорежимность

оба слова каждой пары имеют идентичную шенноновскую энтропию. (По сути, на такой же неразличимости основана Гематрия).