

Размышления. Текст

Под текстом будем понимать все, что может быть представлено цепочкой (возможно, бесконечной) символов заданного алфавита. Иными словами, не только глифы обычного письма, но и, например, вокал или живопись или любой процесс, который может быть разложен в цепочку последовательных различных состояний будем считать текстом в некотором алфавите.

Обычно, текст является мультифракталом. Например, в звуковом фильме можно выделить видео- и аудио- слои, в книге - типографскую структуру (главы, параграфы), слой пунктуации и слой собственно текста в узком смысле слова. Семантика текста образует самостоятельный слой, но его интерпретация невозможна вне рамок более общей модели разделяемого тезауруса и не будет здесь рассматриваться.

В технике связи тексту отводится роль переносчика информации. Более точно, ставится задача клонирования (информация неперемещается) сообщения в новую точку, отделенную от исходной пространственно-временным интервалом.

В пересчете на символ текста, информационная энтропия есть ни что иное, как относительная фрактальная плотность, характеризующая коэффициент использования системы кодов. Единичной (реально недостижимой) энтропии соответствует полное устранение избыточности кодирования (энтропийный предел). Иными словами, в модели Шеннона, количество информации - это "сухой остаток" для предельно спрессованного текста.

Сам "отжим" может быть выполнен различными способами. Исторически, первым был Самюэль Морзе, предложивший использовать неравномерное кодирование: наиболее частые символы алфавита кодировались более короткими последовательностями. Эта идея легла в основу методов Клода Шеннона и Роберта Фано (часто объединяемых в один Шеннона-Фано) и систематически использована Давидом Хаффманом, доказавшим оптимальность предложенного им алгоритма.