

Информационные свойства текста

Главная, общая цель создания текста (любого) – сообщение информации. Любой текст заключает в себе какую-либо информацию. Общее количество информации, содержащейся в тексте, – это его информационная насыщенность.

Текстовая информация широко используется в АСУ при выдаче данных о состоянии объекта и значениях параметров технологического процесса.

Специфические англоязычные тексты применяются при составлении программ на языках высокого уровня. В полиграфическом производстве работа с текстовой информацией приобретает особо важное значение, так как текст является объектом переработки при его кодировании, редактировании, наборе, верстке, корректуре. В процессе работы АСПТИ источник и приемник сообщения обладают дискретными свойствами

В процессе работы ПТИ источники и приемники сообщений обладают дискретными свойствами. Они создают и передают сообщения последовательно один символ за другим. Последовательности символов образующие слова и фразы не являются вполне случайными, а подчиняются определенным статистическим закономерностям.

Для текста имеются ряд понятий, так называемых статистических характеристик.

Статистические характеристики

- Энтропия алфавита – это количество информации, приходящееся на один символ. Символы равновероятного алфавита несут максимальную информационную нагрузку

$$H_{\max} = \log_2 N,$$

где N – объем алфавита.

Алфавиты природных языков не являются равновероятными. Например, относительная частота появления отдельных символов русского языка изменяется от 0,175 до 0,002.

Так же, важной статистической характеристикой текста является избыточность. Избыточностью алфавита называется уменьшение информационной нагрузки на один символ вследствие неравной вероятности и взаимозависимости появления его символов.

- Информационная избыточность, характеризующая относительную недогруженность алфавита рассчитывается так:

$$D = \frac{H_{max}(x) - H(x)}{H_{max}(x)} * 100$$

- Избыточность алфавита определяется по формуле:

$$D = \frac{H_{max}(x) - H(x)}{H_{max}(x)}$$

Где $H_{max}(x)$ - максимальная возможная энтропия при равной вероятности появления знаков.

$H(x)$ – реальная энтропия с учетом вероятностей появления отдельных знаков и вероятностей их различных сочетаний.

- Избыточная информация - это повторяющаяся, дублирующая информация. Она отнимает время у потребителя, отрицательно сказывается на эффективности управления. По имеющимся данным, излишним является в среднем около трети всего объёма информации, предназначенной для потребления в управленческих подразделениях разного уровня.

- Установлено, что в целом ряде случаев избыточная информация приносит даже пользу. Причём необходимость избыточности доказана эволюционным развитием. В частности, избыточными являются существующие алфавиты и языки. К примеру, в русском языке избыточность составляет 40 %, во французском - 55 %. Без некоторой избыточности человек не в состоянии длительное время воспринимать информацию. Это потребовало бы от него постоянного напряжённого внимания, что приводит к быстрому утомлению и потере способности к восприятию.

Наличие избыточности обладает качествами:

- Избыточное сообщение требует дополнительных затрат на передачу и увеличение длительности передач.
- Избыточность повышает помехоустойчивость сообщения, т.к. позволяет обнаруживать и исправлять ошибки при их передаче.
- Высокая избыточность естественных языков общения людей обеспечивает надежное распознавание смысла речи при наличии у отдельных людей различных акцентов и дефектов произношения.