

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

Количественные характеристики информации

Классификация мер информации представлена на рис.1.

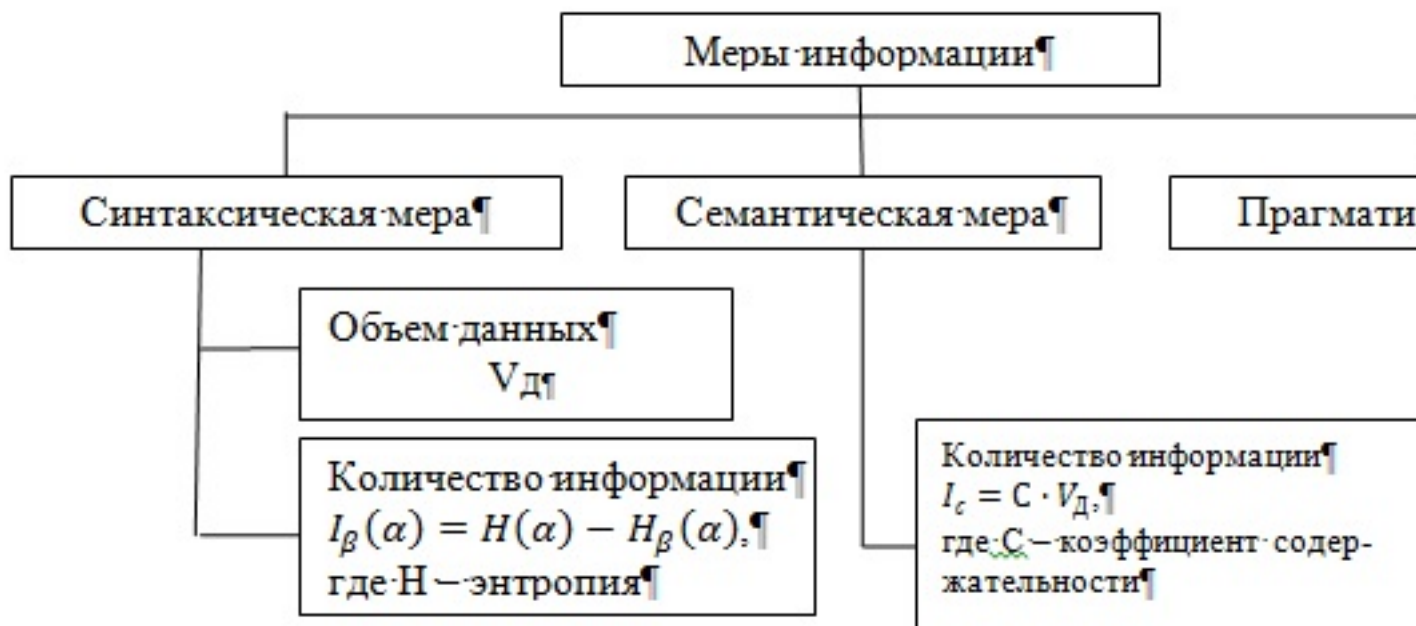


Рис. 1. Классификация мер информации

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

Синтаксическая мера информации оперирует с обезличенной информацией, не выражающей смыслового отношения к объекту. На синтаксическом уровне учитываются тип носителя и способ представления информации, скорость передачи и обработки, размеры кодов представления информации.

Существуют два основных подхода в определении количества информации. Исторически они возникли почти одновременно. В конце 40-х г. XX века один из основоположников кибернетики, американский математик Клод Шеннон развил вероятностный подход к измерению количества информации, а работы по созданию ЭВМ привели к «объемному» подходу.

Объём данных (V_d) понимается в техническом смысле этого слова как информационный объём сообщения или как объём памяти, необходимый для хранения сообщения без каких-либо изменений.

Информационный объём сообщения измеряется в **битах** и равен количеству двоичных цифр (“0” и “1”), которыми закодировано сообщение.

В компьютерной практике слово “бит” используется также как единица измерения объёма памяти. Ячейка памяти размером в 1 бит может находиться в двух состояниях (“включено” и “выключено”) и в неё может быть записана одна двоичная цифра (0 или 1). Понятно, что **бит** – слишком маленькая единица измерения информации, поэтому пользуются кратными ей величинами. Основной единицей измерения информации является **байт**. 1 байт равен 8 битам. В ячейку размером в 1 байт можно поместить 8 двоичных цифр, то есть в одном байте можно хранить $2^8 = 256$

различных чисел. Для измерения ещё больших объёмов информации используются следующие величины:

1 Кбайт (один килобайт) = 2^{10} байт = 1024 байта (1 кВ);

1 Мбайт (один мегабайт) = 2^{10} Кбайт = 1024 Кбайта (1 МВ);

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

1 Гбайт (один гигабайт) = 2^{10} Мбайт = 1024 Мбайта (1 GB);

1 Тбайт (один терабайт) = 2^{10} Гбайт = 1024 Гбайта (1 TB);

1 Пбайт (один петабайт) = 2^{10} Тбайт = 1024 Тбайта (1PB);

1 Эбайт (один эксабайт) = 2^{10} Пбайт = 1024 Пбайта (1 EB);

1 Збайт (один зеттабайт) = 2^{10} Эбайт = 1024 Эбайта (1 ZB);

1 Йбайт (один йоттабайт) = 2^{10} Збайт = 1024 Збайта (1 YB).

Пример 1. При двоичном кодировании текста каждая буква, знак препинания, пробел занимают 1 байт. На странице книги среднего формата примерно 50 строк, в каждой строке около 60 символов, таким образом, полностью заполненная страница имеет объём $50 \times 60 = 3000$ байт ≈ 3 Килобайта. Вся книга среднего формата занимает $\approx 0,5$ Мегабайт. Один номер четырёхстраничной газеты – 150 Килобайт. Если человек говорит по 8 часов в день без перерыва, то за 70 лет он наговорит около 10 Гигабайт информации. Один чёрно-белый кадр (при 32 градациях яркости каждой точки) содержит примерно 300Кб информации, цветной кадр содержит уже около 1Мб информации. Телевизионный фильм продолжительностью 1,5 часа с частотой 25 кадров в секунду - 135 Гб.

При вероятностном подходе количество информации I на синтаксическом уровне определяется через понятие энтропии системы.

Пусть до получения информации потребитель имеет некоторые предварительные (априорные) сведения о системе α . Мерой его неосведомленности о системе является

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

функция $H(\alpha)$, которая в то же время служит и мерой неопределенности состояния системы.

После получения некоторого сообщения β получатель приобрел некоторую дополнительную информацию $I_\beta(\alpha)$, уменьшившую его априорную неосведомленность так, что неопределенность состояния системы после получения сообщения β стала $H_\beta(\alpha)$.

Тогда количество информации $I_\beta(\alpha)$ о системе, полученной в сообщении β , определится как

$$I_\beta(\alpha) = H(\alpha) - H_\beta(\alpha).$$

т.е. количество информации измеряется изменением (уменьшением) неопределенности состояния системы. Если конечная неопределенность $H_\beta(\alpha)$ обратится в нуль, то первоначальное неполное знание заменится полным знанием и количество информации будет определяться как $I_\beta(\alpha) = H(\alpha)$.

Иными словами,

энтропия системы $H(\alpha)$

может рассматриваться как мера недостающей информации.

Энтропия системы $H(\alpha)$, имеющая N возможных состояний, согласно формуле Шеннона, равна:

где p_i – вероятность того, что система находится в i -м состоянии. Для случая, когда все состояния системы равновероятны, т.е. их вероятности равны $1/N$, ее энтропия определяется соотношением:

Пример 2. Часто информация кодируется числовыми кодами в той или иной системе счисления, особенно это актуально при представлении информации в компьютере.

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

Естественно, что одно и то же количество разрядов в разных системах счисления может передавать разное число состояний отображаемого объекта, что можно представить в виде соотношения

$N = m^n$, где N – число всевозможных отображаемых состояний;

m – основание системы счисления (разнообразие символов, применяемых в алфавите);

n – число разрядов (символов) в сообщении.

Допустим, что по каналу связи передается n -разрядное сообщение, использующее m различных символов. Так как количество всевозможных кодовых комбинаций будет $N = m^n$, то при равновероятности появления любой из них количество информации, приобретенной абонентом в результате получения сообщения, будет определяться по формуле Хартли:

$$I = \log N = n \log m$$

Если в качестве основания логарифма принять m , то $I = n$. В данном случае количество информации (при условии полного априорного незнания абонентом содержания сообщения) будет равно объему данных

$$I = V$$

d , полученных по каналу связи.

Наиболее часто используются двоичные и десятичные логарифмы. Единицами измерения в этих случаях будут соответственно **бит** и **дит**.

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

Семантическая мера информации

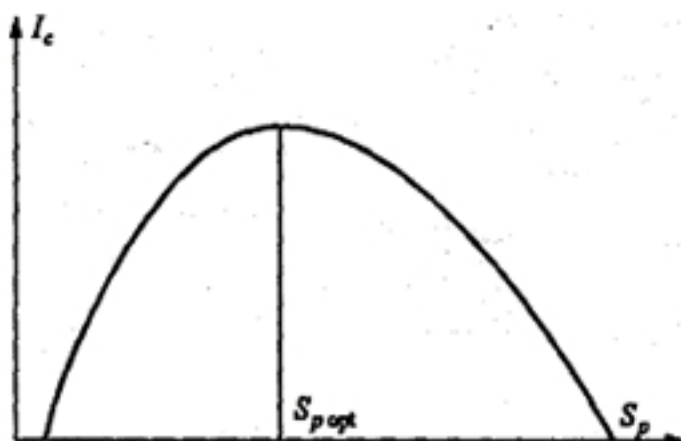
Для измерения смыслового содержания информации, т.е. ее количества на семантическом уровне, наибольшее признание получила тезаурусная мера, которая связывает семантические свойства информации со способностью пользователя принимать поступившее сообщение. Для этого используется понятие «**тезаурус пользователя**»

Тезаурус – это совокупность сведений, которыми располагает пользователь или система.

В зависимости от соотношений между смысловым содержанием информации S и тезаурусом пользователя

S
 p
изменяется количество семантической информации

I
 c
,
воспринимаемой пользователем и включаемой им в дальнейшем в свой тезаурус. Характер зависимости количества семантической информации, воспринимаемой потребителем, от его тезауруса показан на рис. 2.



2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

Рис. 2. Характер зависимости количества семантической информации от его тезауруса

Рассмотрим два предельных случая, когда количество семантической информации I_c равно 0:

- при пользователь не воспринимает и не понимает поступающую информацию;
- при пользователь все знает, и поступающая информация ему не нужна.

Максимальное количество семантической информации I_c потребитель приобретает при согласовании ее смыслового содержания

S
со своим тезаурусом
 S

$$\rho(S) \\ \rho = S$$

ρ_{opt}
)

когда поступающая информация понятна пользователю и несет ему ранее не известные (отсутствующие в его тезаурусе) сведения. Следовательно, количество семантической информации в сообщении, т.е. количество новых знаний, получаемых пользователем, является величиной относительной. Одно и то же сообщение может иметь смысловое содержание для компетентного пользователя и быть бессмысленным для пользователя некомпетентного.

Относительной мерой количества семантической информации может служить коэффициент содержательности C , который определяется как отношение количества семантической информации к ее объему: .

Прагматическая мера информации (аксиологический подход)

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

Эта мера определяет полезность информации (ценность) для достижения пользователем поставленной цели. Эта мера также является величиной относительной, обусловленной особенностями использования этой информации в той или иной системе.

Ценность информации целесообразно измерять в тех же самых единицах (или близких к ним), в которых измеряется целевая функция.

Представим для сопоставления введённые меры информации в таб. 1.

Таб. 1. Меры информации

2. Меры и единицы представления, измерения и хранения информации

Автор: Александр
25.08.2014 16:30

Мера информации	Единицы измерения	Примеры (для компьютерной области)
Синтаксическая: - шенноновский подход; - компьютерный (объемный) подход	Степень уменьшения неопределенности. Единицы представления информации	Вероятность события. Бит, байт, Кбайт и т.д.
Семантическая	Тезаурус. Экономические показатели	Пакет прикладных программ, персональный компьютер, компьютерные сети и т.д. Рентабельность, производительность, коэффициент амортизации и т.д.
Прагматическая	Ценность использования	Емкость памяти, производительность компьютера, скорость передачи данных и т.д. Денежное выражение. Время обработки информации и принятия решений