**Лабораторная работа 1**

**Тема 1. Общие понятия информационных процессов. Основные понятия теории систем**

**Цель занятия:** Получить навыки исследования информационных объектов и определения количество информации в дискретных сообщениях.

**Задачи:**  
1. Провести анализ основных существующих мер информации, применяемых в информационных системах.

2.Подсчитать среднее количество информации, приходящееся на один символ (энтропию) источника дискретных сообщений в случаях

3. Разработать сводную таблицу сравнительных характеристик информационных мер на основе качественного подхода к информации.  
4. Представить процесс передачи информации на примере системы, согласно индивидуально выбранному варианту.

**Индивидуальное задание:**

1 кейс «Едадил» система информирования

**Результаты изучения понятий «энтропия» и «информационные меры»:**

Общая мера неопределенности – **энтропия** - характеризуется некоторой математической зависимостью от совокупности этих вероятностей. Количество информации в сообщении определяется тем, насколько уменьшается эта мера после получения сообщения.

**Количеством информации** называют числовую характеристику сигнала, не зависящую от его формы и содержания, и характеризующую неопределенность, которая исчезнет после получения сообщения в виде данного сигнала.

Для абсолютно достоверного события (событие обязательно произойдёт, поэтому его вероятность равна 1) количество информации в сообщении о нем равно 0. Чем неожиданнее событие, тем больше информации он несет.

**Метрики в информационных системах**

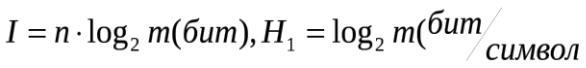
***1.* Комбинаторно*-*вероятностная мера разнообразия заданного множества объектов (мера неопределенности опытов**) Р.Хартли

Количество информации, содержащееся в дискретном сообщении (I), можно найти из простого соотношения:

I=n⋅H,

где n ― число символов в сообщении, H ― энтропия источника сообщений, то есть среднее количество информации, приходящееся на один символ.

В случае, если символы источника сообщения появляются равновероятно и взаимно независимо, то для подсчета энтропии такого рода сообщений используют формулу Хартли:

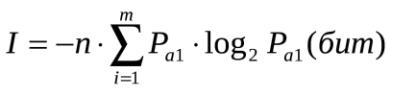
),  
где m ― объем алфавита источника дискретных сообщений.

Характеризуется данная мера тем, что:  
– математическая строгость определения обеспечена за счет абстрагирования от семантических и прагматических свойств (качества) информации;  
– проста и удобна в ряде практических задач связи и управлениях;  
– не учитывает различие между характером имеющихся исходов (почти невероятному исходу придается такое же значение, как и достаточно правдоподобному);

– не отражает случайного характера формирования информационных массивов (ИМ) в задачах анализа, синтеза и управления информационных системами.

***2. Вероятностно-статистическая усредненная мера неопределенности заданного множества объектов*** (статистической «редкости» символов в сообщениях, информационных массивах (ИМ)) Н. Винера и К. Шеннона (селективная энтропия)

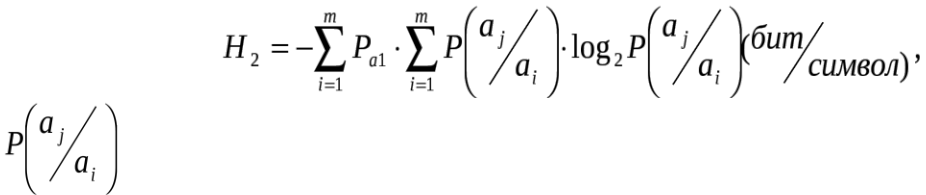
Если символы источника сообщения генерируются с различными вероятностями, но взаимно независимы, то используют формулу Шеннона:,

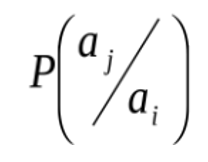
;

;

где Ра1 ― вероятность появления символа a1.

В случае же не равновероятного появления символов источника сообщения и наличия статистических зависимостей между соседними символами энтропию такого источника можно определить с помощью формулы Шеннона с условными вероятностями:



где  ― условная вероятность появления символа aj после символа ai

Название единицы измерения зависит от значения a основания логарифма .

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Основание **а** | **2** | **е** | **3** | **10** | **256** |
| Название единицы измерения | бит(двед) | нат | трит | дит(Хартли) | байт |

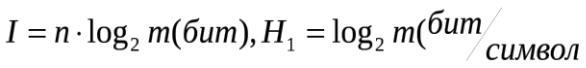
Таким образом, информацию можно измерять длиной сообщения в битах. Такой способ ничего не говорит об информативности сообщения, но характеризует объём работы системы связи при передаче. Если же в задаче необходимо учитывать информативность, то следует пользоваться энтропийным подходом к измерению информации. При этом нужно уточнить, о каком множестве событий будет сообщаться, каковы их вероятности.

**Среднее количество информации, приходящееся на один символ (энтропию) источника дискретных сообщений в случаях:**

В качестве дискретного источника сообщений взять источник с объемом алфавита m= 34 (аналогичный по объему алфавита тексту на русском языке: 33 буквы и пробел), а его статистические характеристики смоделировать с помощью генератора случайных чисел.

а) равновероятного и взаимно независимого появления символов;

В данном случае, для подсчета энтропии мы используем формулу Хартли:

),  
где m ― объем алфавита источника дискретных сообщений.

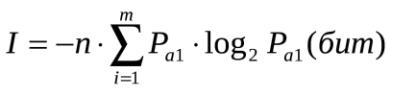
В нашем случае объём алфавита источника сообщения равен m = 34.

Значит энтропия равна:

H = = 5,087 бит/символ

б) Не равновероятного и взаимно независимого появления символов;

Если символы источника сообщения генерируются с различными вероятностями, но взаимно независимы, то используют формулу Шеннона:,

;

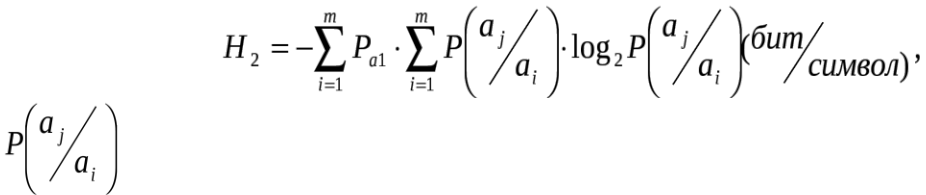
;

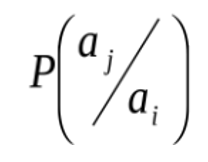
где Ра1 ― вероятность появления символа a1.

Доделать

в) При не равновероятном появлении символов и наличии статистических связей̆ между соседними символами.

В случае не равновероятного появления символов источника сообщения и наличия статистических зависимостей между соседними символами энтропию такого источника можно определить с помощью формулы Шеннона с условными вероятностями:

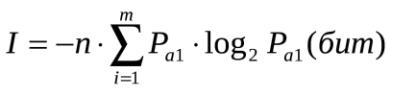


где  ― условная вероятность появления символа aj после символа ai

Доделать

д) Подсчитать количество информации в сообщении, представляющим собой вашу фамилию, имя и отчество, считая, что символы сообщения появляются не равновероятно и независимо. Закон распределения символов найти путем анализа участка любого текста на русском языке длиной не менее 300 символов.

В данном случае нужно использовать формулу Шеннона:

;

;

где Ра1 ― вероятность появления символа a1.

Доделать

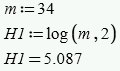
**Программа в электронном варианте (продвинутый уровень):**

m: = 34 ― задание объема алфавита (m);  
i: = 1, 2, ..., m ― i - порядковый номер символа алфавита;

r(i) = rnd (1) ― генерирование 34 случайных чисел в интервале от 0 до 1;

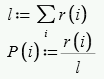
a) Используя формулу Хартли, найти энтропию указанного источника дискретных сообщений (Н1).

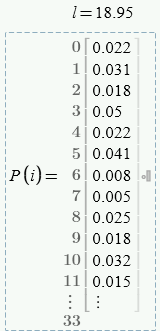
Формула Хартли: 



b) Смоделировать закон распределения символов дискретного источника сообщений, используя оператор rnd (A), который генерирует случайные числа из диапазона [0, A] по следующей программе:

m: = 34 ― задание объема алфавита (m);  
i: = 1, 2, ..., m ― i - порядковый номер символа алфавита;  
r(i) = rnd (1) ― генерирование 34 случайных чисел в интервале от 0 до 1;   
― нахождение суммы всех r(i);   
― P(i) – вероятность появления i-го символа (ai).



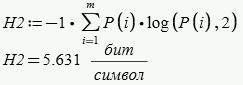


Проверить правильность вычислений, найдя сумму всех P(i) при i= 1,2,...,m.



Используя формулу Шеннона, определить энтропию смоделированного источника дискретных сообщений (Н2).

Формула Шеннона:   
где Ра1 ― вероятность появления символа a1.

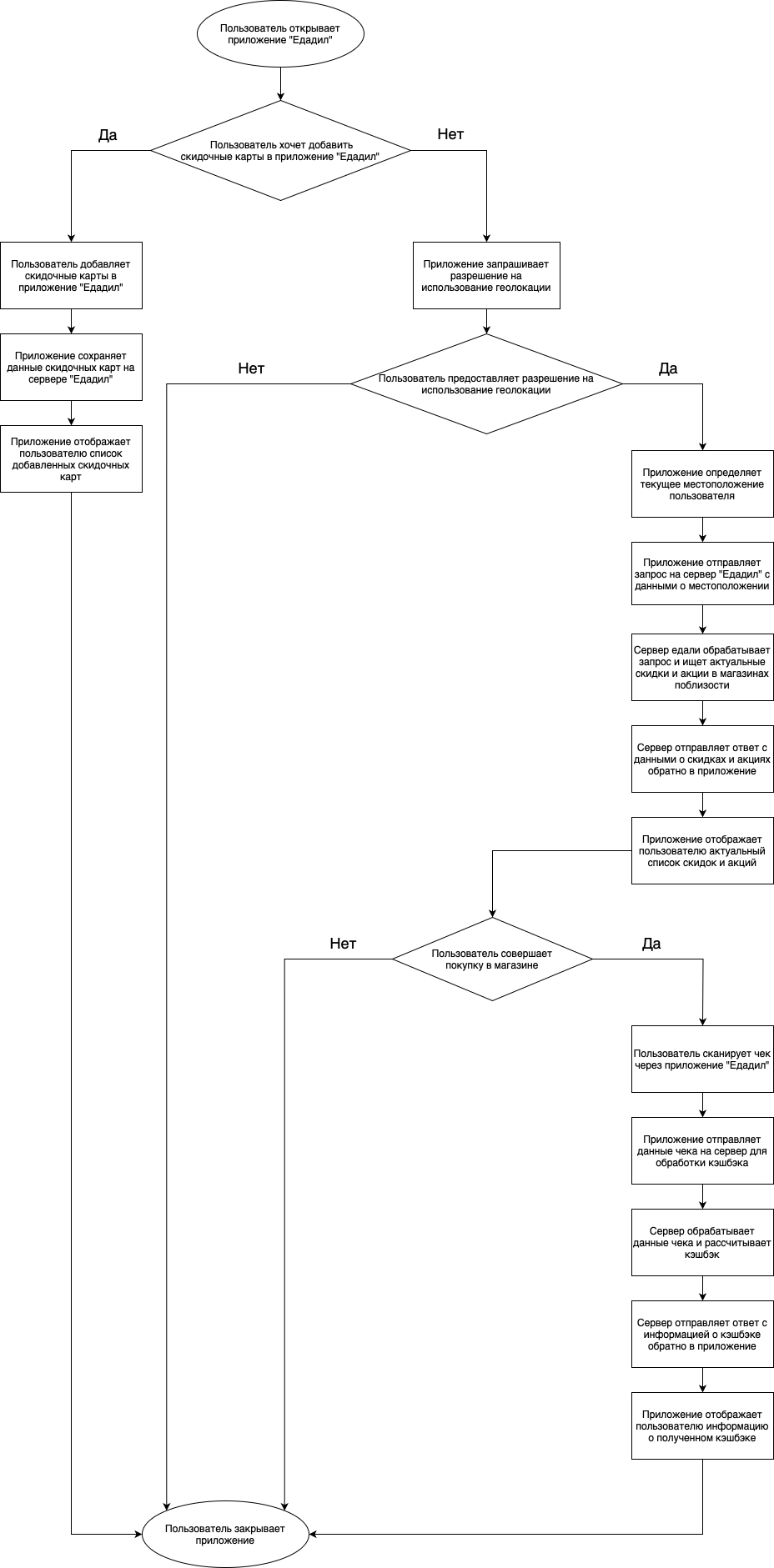


**Сводная таблица сравнительных характеристик информационных мер:**

|  |  |  |  |
| --- | --- | --- | --- |
| Характеристика | Хартли | Шеннона | Шеннона с условными вероятностями |
| Формула |  |  |  |
| Применение | В случае, если символы источника сообщения появляются равновероятно и взаимно независимо | Если символы источника сообщения генерируются с различными вероятностями, но взаимно независимы | В случае не равновероятного появления символов источника сообщения и наличия статистических зависимостей между соседними символами |
| Преимущества | Проста и удобна в ряде практических задач связи и управлениях | Позволяет количественно оценить уровень неопределенности или информации в случайной величине | Позволяет учитывать зависимость между переменными, что делает её более информативной в контексте многомерных распределений. |
| Недостатки | Не учитывает различие между характером имеющихся исходов (почти невероятному исходу придается такое же значение, как и достаточно правдоподобному),  а также не отражает случайного характера формирования информационных массивов (ИМ) в задачах анализа, синтеза и управления информационных системами. | Не учитывает взаимосвязи между случайными величинами, что может быть критично в некоторых приложениях. | Могут потребоваться дополнительные данные и вычисления, особенно для больших наборов данных, а также необходимость наличия достаточного объема данных для адекватной оценки условных вероятностей. |

**Алгоритм процесса передачи информации на примере системы, согласно индивидуально выбранному варианту:**

Вариант: 1 кейс «Едадил» система информирования



**Программная реализация подсчета энтропии источника сообщений:**

ПО имеет в себе 2 заготовленных алфавита — для удобства (Английский и Русский) каждый из алфавитов включает в себя помимо букв еще и пробел.

При запуске программа просит пользователя ввести свой язык и затем ФИО, в качестве сообщения, после она вычисляет энтропию по формуле Хартли, затем генерирует для каждого символа выбранного алфавита случайную вероятность его появления (Сумма всех вероятностей равняется единице).

Далее на основе сгенерированных вероятностей программа вычисляет Энтропию по Шеннону (без зависимостей), после программа генерирует вероятности появления пар символов и вычисляет энтропию по Шеннону (с зависимостями), после программа подсчитывает Количество информации, занимаемое Сообщением в этих трех случаях и выводит все имеющиеся данные в терминал, так-как вероятности имеют достаточно много знаков поле запятой и из за этого вывод тяжело читать — они грубо округляются в выводе на терминал, но в вычислениях используются не округленные величины, поэтому на финальный результат это никак не сказывается.

К отчету будет приложен zip архив содержащий в себе сам код, вывод в формате html (открывается в браузере), этот отчет и скриншоты програмного кода и вывода в терминал.

**Вывод:** в результате выполнения лабораторной изучили теорию и отработали на практике навыки исследования информационных объектов и определения количество информации в дискретных сообщениях. Изучили понятия «энтропия» и «информационные меры», посчитали значения энтропии в различных случаях. Согласно индивидуальному заданию изучили алгоритм процесса передачи информации на примере системы информирования «Едадил» и составили блок схему передачи данных при использовании приложения.