

Балакшин П.В., Соснин В.В. Информатика. – СПб: Университет ИТМО, 2018. – 122 с.

TO REVIEW AND REWRITE!!!! В пособии излагаются основные понятия, необходимые для более глубокого изучения компьютерной техники и систем в будущем. Рассматриваются основные принципы построения, функционирования и организации памяти ЭВМ. Предлагается изучить несколько пакетов, в том числе систему компьютерной верстки TeX, и некоторые варианты распознавания того, что данные были переданы с ошибкой.

Рекомендовано к печати Ученым советом факультета программной инженерии и компьютерной техники.



Университет ИТМО – ведущий вуз России в области информационных и фотонных технологий, один из немногих российских вузов, получивших в 2009 году статус национального исследовательского университета. С 2013 года Университет ИТМО – участник программы повышения конкурентоспособности российских университетов среди ведущих мировых научно-образовательных центров, известной как проект «5 в 100». Цель Университета ИТМО – становление исследовательского университета мирового уровня, предпринимательского по типу, ориентированного на интернационализацию всех направлений деятельности.

©Университет ИТМО, 2018
©Балакшин П.В., Соснин В.В., 2018

Оглавление

0.1	О курсе	2
0.2	Терминология информатики	3
0.3	Терминология теории информации	4
0.4	Признаки классификации информации	6
0.5	Измерение количества информации	7
0.5.1	Мера Хартли	9
0.5.2	Мера Шеннона	10
0.6	Методы получения информации	12
0.6.1	Эмпирические методы	12
0.6.2	Теоретические методы	12
0.6.3	Эмпирико-теоретические методы	13

0.1 О курсе

Цель данного методического пособия состоит в изучении общих принципов работы компьютера и получении навыков работы с рядом пакетов. Студентам предлагается рассмотреть для получения базовых знаний и умений по работе с компьютером такие темы как:

- основы теории информации;
- сжатие компьютерных данных;
- помехоустойчивое кодирование;
- архитектура ЭВМ;
- организация компьютерных сетей;
- работа с офисными пакетами;
- программное обеспечение профессионального программиста.

Наряду с этими темами, авторы этого пособия предоставляют возможность более детально ознакомиться с рядом современных пакетов, в их числе такие широко известные продукты Microsoft, как Microsoft Office Word и Microsoft Excel, и система компьютерной верстки TeX, имеющая свой собственный язык разметки. Освоение указанных пакетов позволит студентам получить полезные навыки по подготовке презентаций, научно-технических отчетов о результатах выполненной работы, в оформлении результатов исследований в виде статей и докладов на научно-технических конференциях. Также авторы пособия продемонстрируют некоторые методики использования программных средств для решения практических задач.

0.2 Терминология информатики

Начать изучение информатики невозможно, не разобравшись в точном значении термина «информатика». Однако, до сих пор в мировой научной общественности не сложилось четкого понимания этого термина. Рассмотрим одно из популярных определений:

Информатика – дисциплина, изучающая свойства и структуру информации, закономерности ее создания, преобразования, накопления, передачи и использования. За рубежом сложилась чуть более узкая трактовка термина информатика. Там под этим понимают пересечение сразу трех областей науки – это информационные технологии, теория информации и computer science. Всё обозначенное выше подходит под определение самого курса "Информатика".

Изучая некоторую науку важно представлять основные даты, вехи её развития:

- 1956(57) – появление термина «информатика» (*нем.* Informatik, Штейнбух).
- 1968 – первое упоминание в СССР (информология, Харкевич).
- 197X – информатика стала отдельной наукой.
- 4 декабря – день российской информатики.

0.3 Терминология теории информации

Рассмотрим некоторые терминологические тонкости. В быденном языке, слова «информация» и «данные» считаются синонимами. Они, как правило, употребляются взаимозаменяемо. И так обстоит дело в информатике и в целом, в компьютерных науках.

Понятие "*информация*" имеет различные трактовки в различных предметных областях. Например, *информация* может пониматься как:

- сигналы для управления, приспособления рассматриваемой системы (в кибернетике);
- мера хаоса в рассматриваемой системе (в физике);
- вероятность выбора в рассматриваемой системе (в теории вероятностей);
- мера разнообразия в рассматриваемой системе (в биологии) и др.

Но мы остановимся на понятиях, близких к информатике.

Информация - это некоторая упорядоченная последовательность сообщений, отражающих, передающих и увеличивающих наши знания.

Информация - это сведения об окружающем мире (объекте, процессе, явлении, событии), которые являются объектом преобразования (включая хранение, передачу и т.д.) и используются для выработки поведения, для принятия решения, для управления или для обучения.

Информация - это новые сведения, подлежащие передаче, хранению и обработке.

Рассмотрим это фундаментальное понятие информатики на основе понятия "*алфавит*" ("алфавитный формальный подход"). Дадим формальное определение *алфавита*.

Алфавит - конечное множество различных знаков (букв), символов, для которых определена операция *конкатенации* (присоединения символа к символу или цепочке символов); с ее помощью по определенным правилам соединения символов и слов можно получать слова (цепочки знаков) и словосочетания (цепочки *слов*) в этом *алфавите* (над этим *алфавитом*).

Знак (буква) - любой элемент алфавита (элемент x алфавита X , где $x \in X$). Понятие знака неразрывно связано с тем, что им обозначается ("со смыслом"), они вместе могут рассматриваться как пара элементов (x, y) , где x - сам знак, а y - обозначаемое этим знаком.

Пример 1:

Примеры *алфавитов*: множество из десяти цифр, множество из знаков русского языка, точка и тире в азбуке Морзе и др. В *алфавите* цифр знак 5 связан с понятием "быть в количестве пяти элементов".

Слово в алфавите (или над алфавитом) - конечная последовательность знаков (букв) алфавита.

Длина $|p|$ некоторого слова p в алфавите (над алфавитом) - число составляющих его букв.

Словарь (словарный запас) - множество различных слов в алфавите (над алфавитом).

В отличие от конечного *алфавита*, словарный запас может быть и бесконечным.

Слова над некоторым заданным *алфавитом* и определяют так называемые *сообщения*.

Пример 2:

Слова над *алфавитом* кириллицы - "Информатика" инто "иини", "и".

Слова над *алфавитом* десятичных цифр и знаков арифметических операций - "1256" "23+78" "35-6+89" "4".

Слова над *алфавитом* азбуки Морзе - ". ." - "— — —".

В *алфавите* должен быть определен порядок следования *букв* (порядок типа "предыдущий элемент - последующий элемент"), то есть любой *ал-*

фавит имеет упорядоченный вид $X = x_1, x_2, \dots, x_n$.

Таким образом, *алфавит* должен позволять решать задачу лексикографического (алфавитного) упорядочивания, или задачу расположения *слов* над этим *алфавитом*, в соответствии с порядком, определенным в *алфавите* (то есть по символам *алфавита*).

0.4 Признаки классификации информации

Рассмотрим две классификации информации. Первая из них - классификация по форме *сообщений* - определенного вида сигналов, символов:

- отношение к источнику или приемнику (входная, выходная и внутренняя);
- отношение к конечному результату (исходная, промежуточная и результирующая);
- актуальность;
- адекватность;
- доступность (открытая, закрытая);
- понятность;
- полнота (достаточная, недостаточная, избыточная);
- достоверность;
- массовость;
- изменчивость (постоянная, переменная, смешанная);
- объективность;
- точность;
- стадия использования (первичная, вторичная);
- ценность.

Вторая классификация - по форме представления информации, способам ее кодирования и хранения:

- графическая;
- звуковая;

- текстовая;
- числовая;
- видеoinформация.

0.5 Измерение количества информации

Любые сообщения измеряются в *байтах, килобайтах, мегабайтах, гигабайтах, терабайтах, петабайтах* и *эксабайтах*, а кодируются, например, в компьютере, с помощью *алфавита* из нулей и единиц, записываются и реализуются в ЭВМ в *битах*.

Приведем основные соотношения между единицами измерения *сообщений*:

- 1 бит (**b**inary digit - двоичное число) = 0 или 1;
- 1 байт = 8 бит;
- 1 килобайт (1Кб) = 2^{13} бит;
- 1 мегабайт (1Мб) = 2^{23} бит;
- 1 гигабайт (1Гб) = 2^{33} бит;
- 1 терабайт (1Тб) = 2^{43} бит;
- 1 петабайт (1Пб) = 2^{53} бит;
- 1 эксабайт (1Эб) = 2^{63} бит.

Теперь нам известно понятие информации, но необходимо еще конкретно знать сколько этой информации. Поэтому есть два важных определения:

Количество информации - число, адекватно характеризующее разнообразие (структурированность, определённости, выбор состояний и т.д.) в оцениваемой системе. Количество информации часто оценивается в битах, причем такая оценка может выражаться и в долях бит (так как речь идет не об измерении или кодировании сообщений).

Мера информации - численная оценка количества информации, которая обычно задана неотрицательной, определенной на множестве событий и являющейся аддитивной функцией (то есть, мера информации объединения событий (множеств) равна сумме мер каждого события). Заметим, что функция меры информации монотонна (при уменьшении или

увеличении вероятности некоторого события количество информации в системе монотонно уменьшается или увеличивается).

Важно: мера вероятности всегда находится в диапазоне от 0 до 1.

Для измерения информации используются различные подходы и методы, например, с использованием меры информации по Р. Хартли и К. Шеннону.

0.5.1 Мера Хартли



Ральф Хартли
1888 – 1970

Пусть известны N состояний системы S (N опытов с различными, равновозможными, последовательными состояниями системы). Если каждое состояние системы закодировать двоичными кодами, то минимальная длина d полученного кода определяется из условия:

$$2^d \geq N \quad \text{или} \quad d \geq \log_2 N$$

Значит, для однозначного описания системы требуется $\log_2 N$ бит. В общем случае количество информации в системе S равно:

$$H_s = \log_k N$$

Единицы измерения количества информации:

- Бит ($k = 2$)
- Трит ($k = 3$)
- Дит (харт) ($k = 10$)
- Нит (нат) ($k = e$)

Примеры использования меры Хартли

Пример 1:

Задание: мальчик загадывает число от 1 до 64. Какое количество вопросов типа "да-нет" понадобится, чтобы гарантированно угадать число?

Решение:

- Первый вопрос: "Загаданное число меньше 32?". Ответ: "Да".
- Второй вопрос: "Загаданное число меньше 16?". Ответ: "Нет".

...

- Шестой вопрос точно приведет к правильному ответу.

Значит, в соответствии с мерой Хартли в загадке мальчика содержится $\log_2 64 = 6$ бит информации ($N = 64$ так как возможно 64 вариантов загаданного числа).

Ответ: 6 бит.

Пример 2:

Задание: мальчик держит за спиной шахматного ферзя и собирается поставить его на произвольную клетку пустой доски. Какое количество информации содержится в его действии?

Решение: шахматная доска имеет размеры 8×8 клеток. Ферзь может быть как белым, так и черным, поэтому количество равновероятных состояний будет равно $8 \times 8 \times 2 = 128$. Получается, количество информации по мере Хартли равно $\log_2 128 = 7$ бит.

Ответ: 7 бит.

Если во множестве $X = x_1, x_2, \dots, x_n$ искать произвольный элемент, то для его нахождения (по Хартли) необходимо иметь не менее $\log_a n$ (единиц) информации.

Уменьшение говорит об уменьшении разнообразия состояний N системы, увеличение говорит об увеличении разнообразия состояний N системы.

Мера Хартли подходит лишь для идеальных, абстрактных систем, так как в реальных системах состояния системы неодинаково осуществимы (неравновероятны).

0.5.2 Мера Шеннона



Клод Шеннон
1916 – 2001

Если состояния системы не равновероятны, используют меру Шеннона. Мера Шеннона оценивает информацию отвлеченно от ее смысла:

$$I = - \sum_{i=1}^N p_i \times \log_2 p_i,$$

где:

I - количество информации, выраженное в битах (в $\log_k p_i$ $k = 2$);

N - число состояний системы;

p_i - вероятность (относительная частота) перехода системы в i -е состояние (вероятность того, что система

находится в состоянии i)

Сумма всех p_i должна быть равна единице.

Если все состояния рассматриваемой системы равновозможны, равновероятны, то есть $p_i = 1/n$, то из *формулы Шеннона* можно получить (как частный случай) *формулу Хартли*:

$$I = \log_2 n.$$

Обозначим величину:

$$f_i = -n \log_2 p_i.$$

Тогда из *формулы К. Шеннона* следует, что количество информации I можно понимать как среднеарифметическое величин f_i , то есть величину f_i можно интерпретировать как *информационное содержание символа алфавита* с индексом i и величиной p_i вероятности появления этого символа в любом сообщении (слове), передающем информацию.

В термодинамике известен так называемый коэффициент Больцмана $k = 1.38 \cdot 10^{-16} (.)$ и выражение (*формула Больцмана*) для энтропии или меры хаоса в термодинамической системе:

$$S = -k \sum_{i=1}^N p_i \times \ln p_i$$

Сравнивая выражения для I и S , можно заключить, что величину I можно понимать как энтропию из-за нехватки информации в системе (о системе).

Формулы энтропии и информации идентичны, но смысл разный. Энтропия априорная характеристика (до передачи), информация – апостериорная (после передачи).

Из этой формулы следуют важные выводы:

- увеличение меры Шеннона свидетельствует об уменьшении энтропии (увеличении порядка) системы;
- уменьшение меры Шеннона свидетельствует об увеличении энтропии (увеличении беспорядка) системы.

Положительная сторона *формулы Шеннона* – ее отвлеченность от смысла информации. Кроме того, в отличие от *формулы Хартли*, она учитывает различность состояний, что делает ее пригодной для практических вычислений. Основная отрицательная сторона *формулы Шеннона* – она не распознает различные состояния системы с одинаковой вероятностью.

Примеры использования меры Шеннона

Пример 1:

Задание: девочка наугад вытаскивает из мешка мяч. Известно, что в мешке всего 8 мячей, из них: 4 красных, 2 синих, 1 зеленый и 1 белый. Какое количество информации содержится в этом событии?

Решение:

- Вероятность вытащить красный мяч равна $4/8 = 0,5$
- Вероятность вытащить синий мяч равна $2/8 = 0,25$

• Вероятность вытащить зеленый мяч равна $1/8 = 0,125$

• Вероятность вытащить белый мяч равна $1/8 = 0,125$

Значит количество информации, выраженное в битах равно: $I = -(0,5 \times \log_2 0,5 + 0,25 \times \log_2 0,25 + 0,125 \times \log_2 0,125 + 0,125 \times \log_2 0,125) = -(-0,5 \times 1 - 0,25 \times 2 - 0,125 \times 3 - 0,125 \times 3) = -(-0,5 - 0,5 - 0,375 - 0,375) = 1,75$ бит.

Ответ: 1,75 бит

0.6 Методы получения информации

Методы получения информации можно разбить на три большие группы:

- *Эмпирические;*
- *Теоретические;*
- *Эмпирико-теоретические.*

Кратко рассмотрим и охарактеризуем все три метода по отдельности.

0.6.1 Эмпирические методы

Эмпирические методы или методы получения эмпирических данных.

1. *Наблюдение* – сбор первичной информации об объекте, процессе, явлении.
2. *Сравнение* – обнаружение и соотнесение общего и различного.
3. *Измерение* – поиск с помощью измерительных приборов эмпирических фактов.
4. *Эксперимент* – преобразование, рассмотрение объекта, процесса, явления с целью выявления каких-то новых свойств.

Кроме классических форм их реализации, в последнее время используются опрос, интервью, тестирование и другие.

0.6.2 Теоретические методы

Теоретические методы или методы построения различных теорий.

1. *Восхождение от абстрактного к конкретному* – получение знаний о целом или о его частях на основе знаний об абстрактных проявлениях в сознании, в мышлении.

2. *Идеализация* – получение знаний о целом или его частях путем представления в мышлении целого или частей, не существующих в действительности.
3. *Формализация* – получение знаний о целом или его частях с помощью языков искусственного происхождения (формальное описание, представление).
4. *Аксиоматизация* – получение знаний о целом или его частях с помощью некоторых аксиом (не доказываемых в данной теории утверждений) и правил получения из них (и из ранее полученных утверждений) новых верных утверждений.
5. *Виртуализация* – получение знаний о целом или его частях с помощью искусственной среды, ситуации.

0.6.3 Эмпирико-теоретические методы

Эмпирико-теоретические методы (смешанные) или методы построения теорий на основе полученных эмпирических данных об объекте, процессе, явлении.

1. *Абстрагирование* – выделение наиболее важных для исследования свойств, сторон исследуемого объекта, процесса, явления и игнорирование несущественных и второстепенных.
2. *Анализ* – разведение целого на части с целью выявления их связей.
3. *Декомпозиция* – разведение целого на части с сохранением их связей с окружением.
4. *Синтез* – соединение частей в целое с целью выявления их взаимосвязей.
5. *Композиция* – соединение частей целого с сохранением их взаимосвязей с окружением.
6. *Индукция* – получение знания о целом по знаниям о частях.
7. *Дедукция* – получение знания о частях по знаниям о целом.
8. *Эвристики, использование эвристических процедур* – получение знания о целом по знаниям о частях и по наблюдениям, опыту, интуиции, предвидению.

9. *Моделирование (простое моделирование)*, использование приборов – получение знания о целом или о его частях с помощью модели или приборов.
10. *Исторический метод* – поиск знаний с использованием предыстории, реально существовавшей или же мыслимой.
11. *Логический метод* – поиск знаний путем воспроизведения частей, связей или элементов в мышлении.
12. *Макетирование* – получение информации по макету, представлению частей в упрощенном, но целостном виде.
13. *Актуализация* – получение информации с помощью перевода целого или его частей (а следовательно, и целого) из статического состояния в динамическое состояние.
14. *Визуализация* – получение информации с помощью наглядного или визуального представления состояний объекта, процесса, явления.

Кроме указанных классических форм реализации теоретико-эмпирических методов часто используются и мониторинг (система наблюдений и анализа состояний), деловые игры и ситуации, экспертные оценки (экспертное оценивание), имитация (подражание) и другие формы.

Пример :

Для построения модели планирования и управления производством в рамках страны, региона или крупной отрасли нужно решить следующие проблемы:

1. определить структурные связи, уровни управления и принятия решений, ресурсы; при этом чаще используются методы наблюдения, сравнения, измерения, эксперимента, анализа и синтеза, дедукции и индукции, эвристический, исторический и логический методы, макетирование и др.;
2. определить гипотезы, цели, возможные проблемы планирования; наиболее используемые методы – наблюдение, сравнение, эксперимент, абстрагирование, анализ, синтез, дедукция, индукция, эвристический, исторический, логический и др.;
3. конструирование эмпирических моделей; наиболее используемые методы – абстрагирование, анализ, синтез, индукция, дедукция, формализация, идеализация и др.;

4. поиск решения проблемы планирования и просчет различных вариантов, директив планирования, поиск оптимального решения; используемые чаще методы – измерение, сравнение, эксперимент, анализ, синтез, индукция, дедукция, актуализация, макетирование, визуализация, виртуализация и др.