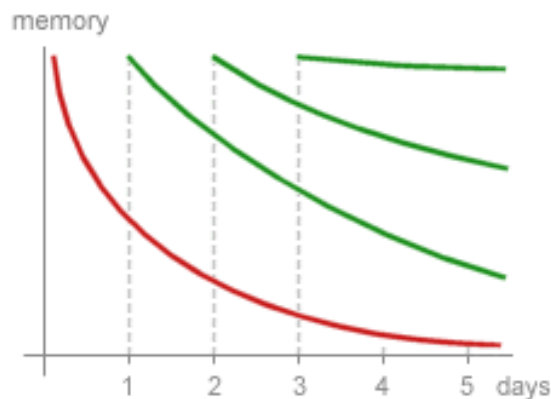




Проектирование и эксплуатация информационных систем в медиаиндустрии

Выломова Екатерина Алексеевна
e-mail: evylotova@gmail.com

0. Кривая забывания



Опыт Эббингауза с запоминанием
бессмысленных слогов:

1ый час – 60% информации забыто
через 10 часов – 35% остается
6 дней – месяц – 20%

Запоминание:

2 дня: 1 – сразу после прочтения; 2 – спустя 20 минут после 1-го прочтения; 3 – через 8 часов после 2-го; 4 – через 24 часа после 3-го.

Надолго: 1- сразу после запоминания; 2 – через 20-30 минут после 1-го повторения; 3 – через день после 2-го; 4 – через 2-3 недели после 3-го; 5 – через 2 -3 месяца после 4-го.

Осмысленное запоминание в **9раз** быстрее механического заучивания

Эффект края – лучше всего запоминается материал, находящийся в начале и в конце.

Кривая забывания

0. Бритва Оккама

«Не следует множить сущее без необходимости»



или

«Не следует привлекать новые сущности без самой крайней на то необходимости»



I. Базовые понятия

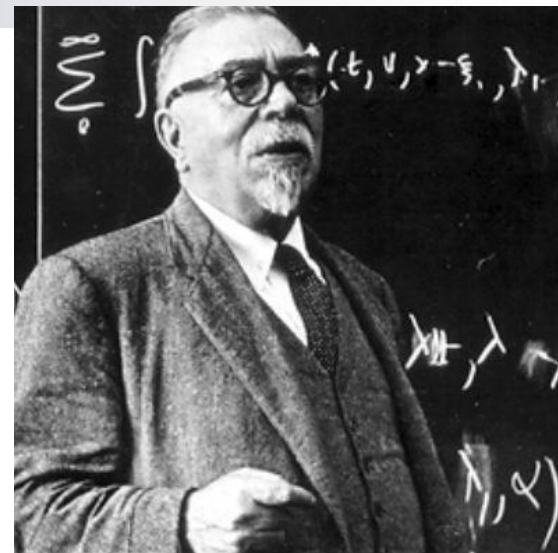
- Кибернетика
- Форма адекватности
- Мера информации
- Качество информации
- Информационная модель

I. Кибернетика

др. греч. – «искусство управления»

Наука об общих законах получения, хранения, передачи и переработки информации.

Основатель: Норберт Винер



The best material model of a cat is another, or preferably the same, cat. // Philosophy of Science (1945) (with A. Rosenblueth)

Включает в себя: теорию автоматов,
теорию алгоритмов, теория принятия решений,
исследование операций, распознавание
образов

Объект изучения: все управляемые системы

I. Кибернетика

Кибернетика — наука об оптимальном управлении сложными динамическими системами.

Оптимальное управление — это перевод системы в новое состояние с выполнением некоторого критерия оптимальности, например, минимизации затрат времени, труда, веществ или энергии.

Принцип управления заключается в том, что движение и действие больших масс или передача и преобразование больших количеств энергии направляется и контролируется при помощи небольших количеств энергии, несущих информацию.



I. Теория информации

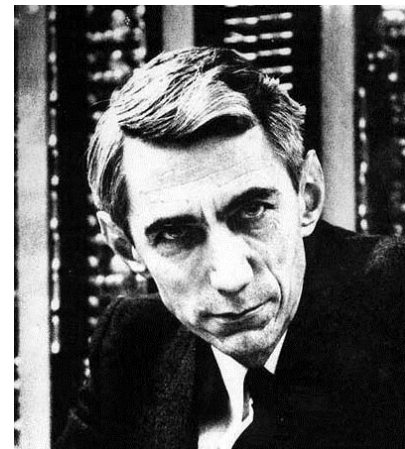
Информация - сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности, неполноты знаний.

Теория информации — раздел прикладной математики, аксиоматически определяющий понятие информации, её свойства и устанавливающий предельные соотношения для систем передачи данных.

Основатель: Клод Шеннон

Базовые понятия теории информации:

1. Информация
2. Канал связи
3. Шум
4. Кодирование
5. Данные
6. Адекватность информации



I. Формы адекватности

Адекватность информации - это определенный уровень соответствия создаваемого с помощью полученной информации образа реальному объекту, процессу, явлению и т.п.

- Синтаксическая -> Синтаксическая мера

Формально-структурные характеристики, без учета смысла

Объем данных и энтропия

- Семантическая

Учет смысла. Вводится через учет тезауруса пользователя и соответствия новой информации имеющейся

- Прагматическая

Отношение информации и ее потребителя, соответствие цели управления

I. Классификация мер

Два основных параметра:

- Количество информации I
- Объем данных V_d



I. Синтаксическая мера

- 1865 г. : Рудольф Клазиус вводит понятие энтропии
- 1921 г.: Рональд Фишер вводит термин «информация»
- 1848 г. : Клод Шеннон разрабатывает теорию связи: формулы вычисления количества информации и меры неопределенности в системе

I. Энтропия и информация

Энтропия (формула Шеннона): $H(\alpha) = -\sum_{i=1}^N P_i \log P_i$

где P_i - вероятность того, что система находится в i -м состоянии.

Под энтропией понимается мера неопределенности распределения

События равновероятны:

$$H(\alpha) = -\sum_{i=1}^N \frac{1}{N} \log \frac{1}{N}$$

Теорема Хартли: Информативность символа m -элементного алфавита равна $\log m$.

Формула Шеннона выражает информативность источника информации с m -символьным алфавитом и данной частотной характеристикой.

Коэффициент(степень) информативности: $Y = \frac{I}{V_d}$, причем $Y < 0 < 1$,

I. Энтропия и информация

1. Равномерное распределение имеет наибольшую энтропию среди всех распределений с данным числом исходов.
2. Информация по Шеннону:

- Дискретный случай $I(X, Y) = \sum_{i,j} p_{ij} \log_2 \frac{p_{ij}}{p_i q_j}$

- Непрерывный случай $I(X, Y) = \iint_{\mathbb{R}^2} p_{XY}(t_1, t_2) \log_2 \frac{p_{XY}(t_1, t_2)}{p_X(t_1)p_Y(t_2)} dt_1 dt_2.$

$$I(X, X) = \sum_i p_i \log_2 \frac{p_i}{p_i p_i} = - \sum_i p_i \log_2 p_i.$$

Энтропия д.с.в. X в теории информации определяется формулой

$$H(X) = H X = I(X, X).$$

I. Энтропия и информация

Свойства меры информации и энтропии:

- 1) $I(X, Y) \geq 0$, $I(X, Y) = 0 \Leftrightarrow X$ и Y независимы;
- 2) $I(X, Y) = I(Y, X)$;
- 3) $HX = 0 \Leftrightarrow X$ — константа;
- 4) $I(X, Y) = HX + HY - H(X, Y)$, где $H(X, Y) = -\sum_{i,j} p_{ij} \log_2 p_{ij}$;
- 5) $I(X, Y) \leq I(X, X)$. Если $I(X, Y) = I(X, X)$, то X — функция от Y .

Задачи:

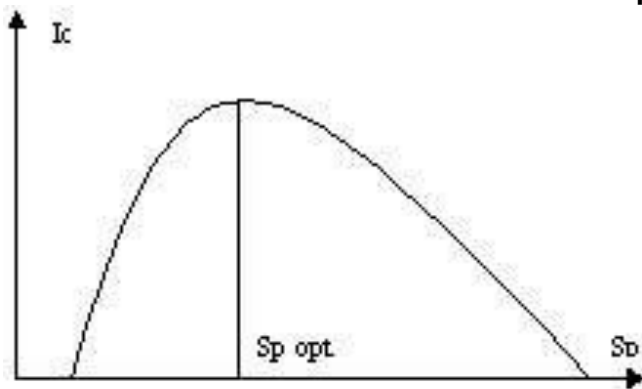
1. Во сколько раз можно сжать файл, представляющий последовательность из трех символов a, b, c , если известно, что частота a равна $\frac{1}{2}$ и после a в половине случаев идет b , а в половине c ?
2. В игре «Поле чудес» вам выпала возможность открыть сразу две буквы слова. Какую из позиций следует выбрать?

I. Семантическая мера

Основная – тезаурусная мера

Тезаурус – совокупность сведений, которыми располагает пользователь или система

Тезаурусная мера связывает семантические свойства информации со способностью пользователя принимать поступившее сообщение.



коэф. содержательности: $C = \frac{I_c}{V_D}$

Зависимость количества семантической информации, воспринимаемой потребителем, от его тезауруса $I_c = f(S_p)$

Введение. Семантическая мера информации

I. Прагматическая мера

Определяет полезность (ценность) информации для достижения пользователем поставленной цели.

Пример: В экономической системе прагматические свойства (ценность) информации можно определить приростом экономического эффекта функционирования, достигнутым благодаря использованию этой информации для управления системой:

$$Inb(g) = \Pi(g/b) - \Pi(g),$$

где $Inb(g)$ - ценность информационного сообщения b для системы управления g ,

$\Pi(g)$ - априорный ожидаемый экономический эффект функционирования системы управления g ,

$\Pi(g/b)$ - ожидаемый эффект функционирования системы g при условии, что для управления будет использована информация, содержащаяся в сообщении b .

Введение. Прагматическая мера информации

I. Качество информации

- Репрезентативность
- Содержательность
- Достаточность
- Полнота
- Доступность
- Актуальность
- Своевременность
- Точность
- Достоверность
- Устойчивость



II. Информационные модели и технологии

Информационная модель — формальная модель ограниченного набора фактов, понятий или инструкций, предназначенная для удовлетворения конкретному требованию.

Информационная технология — совокупность методов, производственных и программно-технологических средств, объединенных в технологическую цепочку, обеспечивающую сбор, хранение, обработку, вывод и распространение информации.

Этапы развития информационных технологий:

1. Становление
2. Активный рост
3. Общедоступность

Репликативный цикл технологии – время, необходимое на копирование технологии конкурентом.

II. Информационная система

Система – ограниченное, взаимно связанное множество, отражающее объективное существование конкретных отдельных взаимосвязанных совокупностей объектов и не содержащее специфических ограничений, присущих частым системам.

Свойства системы:

1. Ограниченность
2. Целостность организации
3. Структурность
4. Взаимосвязь со средой
5. Иерархичность
6. Множественность описаний

Системный подход предполагает:

- Рассмотрение объекта как системы
- Установление состава, структуры и системы
- Выделение внешних связей
- Определение функций и роли системы
- Анализ противоречий структуры и функций системы
- Определение тенденций развития

II. Информационная система

Информационная система - взаимосвязанная совокупность средств, методов и персонала, используемых для хранения, обработки и выдачи информации в интересах достижения поставленной цели.

Характеристики крупных проектов создания ИС:

1. Сложность описания;
2. Множество взаимодействующих компонентов;
3. Ограниченная возможность применения типовых решений;
4. Необходимость интеграции с имеющимися системами;
5. Неоднородная среда выполнения;
6. Разнородность и разобщенность групп разработчиков;
7. Временная протяженность проекта.

II. Процессы и потоки

Типы процессов:

1. Ввод информации
2. Обработка входной информации
3. Вывод информации
4. Процесс обратной связи

Информационный поток — это совокупность циркулирующих в системе, между системой и внешней средой сообщений, необходимых для управления, анализа и контроля операций.

Свойства информационного потока:

1. Источник возникновения;
2. Направление движения;
3. Скорость передачи и приема;
4. Интенсивность потока.

II. Состав и классификация

Типы обеспечивающих подсистем:

1. Информационное обеспечение
2. Техническое обеспечение
3. Математическое и программное обеспечение
4. Организационное обеспечение
5. Правовое обеспечение
6. Программное обеспечение

По степени автоматизации:

- 1. Ручные
- 2. Автоматизированные
- 3. Автоматические

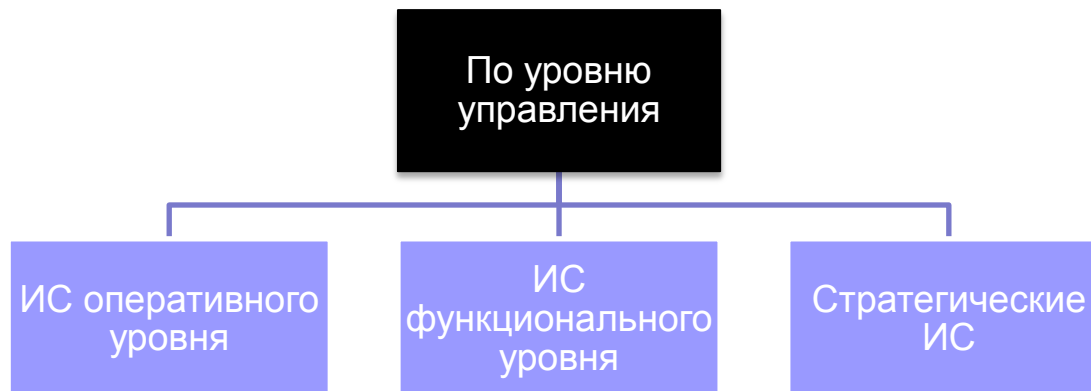
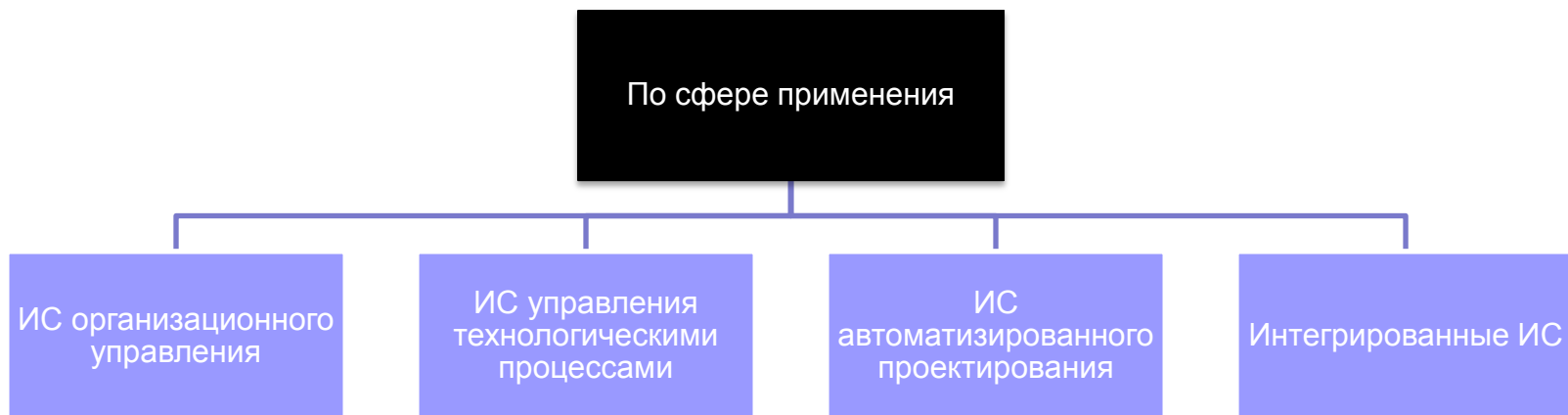
По характеру обработки данных

- 1. Информационно-поисковые
- 2. Информационно-решающие
 - 2.1 Управляющие
 - 2.2 Советующие

По типу данных:

- 1. Фактографические
- 2. Документные

II. Классификация ИС





II. FIN

Следующее занятие:

- Архитектура информационных систем
- Конкретный пример: архитектура и принципы работы поисковых машин