

В оный день, когда над миром новым
Бог склонял лицо Своё, тогда
Солнце останавливали словом,
Словом разрушали города.

И орёл не взмахивал крылами,
Звёзды жались в ужасе к луне,
Если, точно розовое пламя,
Слово проплывало в вышине.

(Николай Гумилёв)

Текстовая информация, инструменты для её компьютерной обработки.

Информатика, 1 курс.
Лекция 5.

В ЭТОЙ ЛЕКЦИИ:

- Механизм восприятия текстовой информации.
- Представление текста в компьютере.
- Классификация программ, работающих с текстовой информацией.
- Способы улучшения восприятия текстовой информации

Структурные элементы текстовой информации



Смысловой единицей в текстовой информации является слово. Однако, как правило, человек воспринимает, запоминает, использует для анализа не слово, а текст в целом.

Представление текста в компьютере

ASCII (англ. American Standard Code for Information Interchange)

Основная таблица ASCII

| | 00 | 10 | 20 | 30 | 40 | 50 | 60 | 70 |
|---|----|----|----|----|----|----|----|----|
| 0 | | ▯ | | 0 | @ | P | ' | ▯ |
| 1 | ▯ | ▯ | ! | 1 | A | Q | a | q |
| 2 | ▯ | ▯ | " | 2 | B | R | b | r |
| 3 | ▯ | !! | # | 3 | C | S | c | s |
| 4 | ▯ | ▯ | \$ | 4 | D | T | d | t |
| 5 | ▯ | § | % | 5 | E | U | e | u |
| 6 | ▯ | = | & | 6 | F | V | f | v |
| 7 | + | ▯ | ' | 7 | G | W | g | w |
| 8 | ▯ | ↑ | < | 8 | H | X | h | x |
| 9 | o | ↓ | > | 9 | I | Y | i | y |
| A | ▯ | + | * | : | J | Z | j | z |
| B | ▯ | ← | + | ; | K | [| k | < |
| C | ▯ | └ | , | < | L | \ | l | : |
| D | ▯ | ↔ | - | = | M |] | m | > |
| E | ▯ | ▯ | . | > | N | ^ | n | ~ |
| F | ▯ | ▯ | / | ? | O | _ | o | ▯ |

Расширенная таблица ASCII (cp866)

| | 80 | 90 | A0 | B0 | C0 | D0 | E0 | F0 |
|---|----|----|----|----|----|----|----|----|
| 0 | А | Р | а | ▯ | └ | └ | Р | ▯ |
| 1 | Б | С | б | ▯ | └ | └ | с | ▯ |
| 2 | В | Т | в | ▯ | └ | └ | т | ▯ |
| 3 | Г | У | г | | └ | └ | у | < |
| 4 | Д | Ф | д | └ | └ | └ | ф | ▯ |
| 5 | Е | Х | е | └ | └ | └ | х | ▯ |
| 6 | Ж | Ц | ж | ▯ | └ | └ | ц | ▯ |
| 7 | З | Ч | з | ▯ | └ | └ | ч | ▯ |
| 8 | И | Ш | и | └ | └ | └ | ш | ▯ |
| 9 | Й | Щ | й | ▯ | └ | └ | щ | ▯ |
| A | К | Ь | к | ▯ | └ | └ | ь | ▯ |
| B | Л | Ы | л | ▯ | └ | ▯ | ы | ▯ |
| C | М | Ъ | м | ▯ | └ | ▯ | ъ | ▯ |
| D | Н | Э | н | ▯ | ▯ | ▯ | э | ▯ |
| E | О | Ю | о | ▯ | ▯ | ▯ | ю | ▯ |
| F | П | Я | п | ▯ | ▯ | ▯ | я | ▯ |

Представление текста в компьютере

Юникод, или Уникод (англ. Unicode) — стандарт кодирования символов, позволяющий представить знаки практически всех письменных языков.

- Плоскость 0 (0000—FFFF): Базовая многоязыковая плоскость
- Плоскость 1 (10000—1FFFF): Дополнительная многоязыковая плоскость
- Плоскость 2 (20000—2FFFF): Дополнительная иероглифическая плоскость
- Плоскость 3 (30000—3FFFF): Третичная иероглифическая плоскость
- Плоскости 4—13 (40000—DFFFF) не используются
- Плоскость 14 (E0000—EFFFF): Дополнительная плоскость особого назначения
- Плоскость 15 (F0000—FFFFF) используется как дополнительная область-А для частного использования
- Плоскость 16 (100000—10FFFF) используется как дополнительная область-В для частного использования

Базовая многоязыковая плоскость

| | | | | | | | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 00 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 0A | 0B | 0C | 0D | 0E | 0F |
| 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 1A | 1B | 1C | 1D | 1E | 1F |
| 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 2A | 2B | 2C | 2D | 2E | 2F |
| 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 3A | 3B | 3C | 3D | 3E | 3F |
| 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 4A | 4B | 4C | 4D | 4E | 4F |
| 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 5A | 5B | 5C | 5D | 5E | 5F |
| 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 6A | 6B | 6C | 6D | 6E | 6F |
| 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 7A | 7B | 7C | 7D | 7E | 7F |
| 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 8A | 8B | 8C | 8D | 8E | 8F |
| 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 9A | 9B | 9C | 9D | 9E | 9F |
| A0 | A1 | A2 | A3 | A4 | A5 | A6 | A7 | A8 | A9 | AA | AB | AC | AD | AE | AF |
| B0 | B1 | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B9 | BA | BB | BC | BD | BE | BF |
| C0 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | CA | CB | CC | CD | CE | CF |
| D0 | D1 | D2 | D3 | D4 | D5 | D6 | D7 | D8 | D9 | DA | DB | DC | DD | DE | DF |
| E0 | E1 | E2 | E3 | E4 | E5 | E6 | E7 | E8 | E9 | EA | EB | EC | ED | EE | EF |
| F0 | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | FA | FB | FC | FD | FE | FF |

Чёрный — расширенный латинский алфавит;

Голубой — лингвистические символы международного фонетического алфавита IPA;

Синий — другие европейские алфавиты;

Оранжевый — письменности Ближнего Востока;

Светло-оранжевый — письменности Африки;

Зелёный — письменности Южной Азии;

Фиолетовый — письменности Юго-восточной Азии;

Красный — письменности Восточной Азии;

Розовый — унифицированные китайско-японско-корейские символы;

Жёлтый — письменности аборигенов Северной Америки;

Пурпурный — символы;

Тёмно-серый — диакритики;

Светло-серый — суррогатные пары UTF-16 и области для частного использования;

Циан — другие знаки;

Белый — не используется.

Программы для работы с текстовой информацией

Вьюверы (от viewer) — просмотрщики, компьютерные программы, предназначенные для просмотра файлов.

Редакторы – программы для создания и корректировки текстов

Специализированные программы – например, системы для машинного перевода, лингвистического анализа, шифрования и дешифровки.

Редакторы текста

```
graph TD; A[Редакторы текста] --> B[Текстовые редакторы  
– ввод и корректировка текста]; A --> C[Текстовые процессоры –  
форматирование текста,  
вставка иллюстраций,  
автоматизация работы с  
текстом]; A --> D[Рабочие среды,  
включающие редактор  
текста (например,  
системы  
программирования)]; A --> E[Издательские системы -  
подготовка оригинал-  
макетов для полиграфии];
```

Текстовые редакторы
– ввод и корректировка
текста

Текстовые процессоры –
форматирование текста,
вставка иллюстраций,
автоматизация работы с
текстом

Рабочие среды,
включающие редактор
текста (например,
системы
программирования)

Издательские системы -
подготовка оригинал-
макетов для полиграфии

| Критерии сравнения | Текстовый редактор | Текстовый процессор | Издательская система | Рабочая среда |
|---|---------------------------|---|---|---|
| Пользователь | Практически любой | Работник офиса, учёный, преподаватель, студент и т.п. | Редактор, верстальщик в издательстве | Пользователь конкретной среды (например, программист) |
| Внутреннее представление | Только коды символов | Текст и «тэги» его форматирования, ссылки на встроенные объекты | Блоки текста и иллюстраций и их характеристики | Текст и «тэги» автоформатирования |
| Наличие форматирования | Отсутствует | Развитое | Акцент на вёрстке – расположение блоков текста. | Автоматическое – в соответствии с видом элементов. |
| Автоматизация создания типовых документов | Отсутствует | Развитая | Развитая | Присутствует |
| Работа с графическими и иными объектами | Нет | Есть | Есть | Нет |

Восприятие текстовой информации

В связи с ликвидацией кафе «Диарея» дёшево продаются оборудование и продукты: холодильник, дровяная печь, картофель, макароны, электровеник, сахар, соль, кастрюли. Адрес: село Непруха, ул. Заовражная, дом 13. Телефон 13-13-13

Чувствуете
разницу?

Тогда
объясните,
в чём она.

В связи с ликвидацией кафе «Диарея»
ДЁШЕВО ПРОДАЮТСЯ

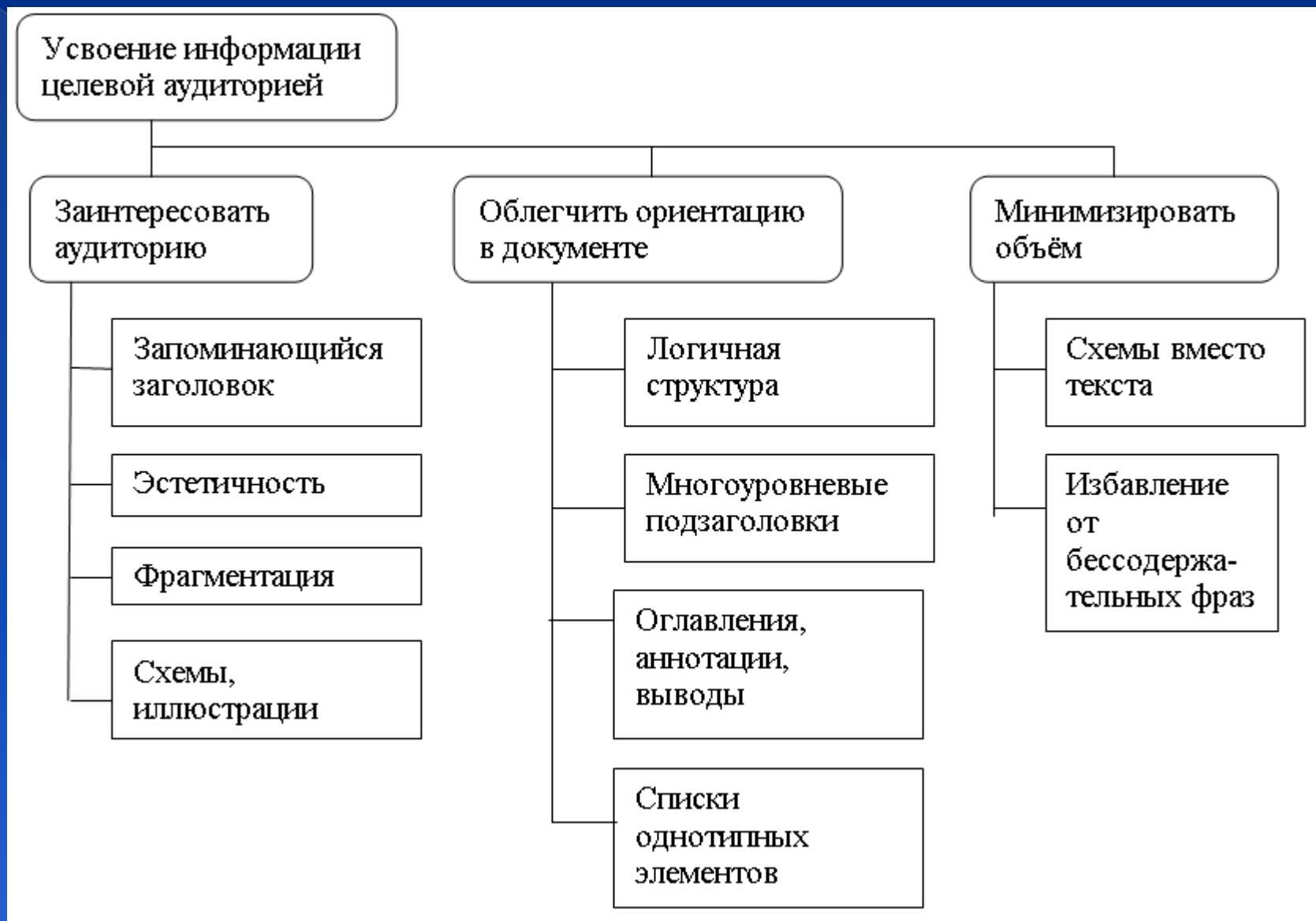
Оборудование:

- холодильник
- дровяная печь
- электровеник
- кастрюли

Продукты:

- картофель
- макароны
- сахар
- соль

Адрес: село Непруха, ул. Заовражная, дом 13.
Телефон 13-13-13



Неструктурированный текст

Беспроblemная сдача экзаменов: шпаргалки или микронаушники ТЕХ? Здесь никто не станет убеждать вас, что лучше никогда не писать шпаргалок и учиться, учиться и еще раз учиться. Если можете учиться, принося свою молодую жизнь на алтарь победы, знаете наизусть всю Большую Советскую энциклопедию, насмешливо глядите на преподавателя с выражением «Я знаю больше чем ты», тогда вы Анатолий Вассерман, зачем вы здесь? Из любопытства? Да ладно. Здесь мы поговорим о шпаргалках, и как правильно подготовиться к их использованию. Для начала, каждый преподаватель должен знать, что вы человек чертовски рассеянный и параноидальный, вы постоянно роняете вещи, прячете от всех, что-то в одежде. Вы всегда закрываете рукой учебники и тетрадки, потому что вам кажется, что за вами подглядывают и списывают, делайте это достаточно часто, чтобы ваше поведение принималось как само собой разумеющееся, тогда пол дела сделано. Параллельно с ведением конспекта старайтесь заносить всю информацию в персональный компьютер, если не сможете сами охватить весь объем, агитируйте других членов группы, мол, зачем раздавать конспекты, если можно просто делиться файлами, попробуйте предложить преподавателю перенести его конспект на компьютер, чтобы им было удобнее пользоваться, ну суть вы поняли. Это поможет отложить немного информации в себя, тем же набором текста ну и на экзамене это вам пригодится при изготовлении шпаргалок. Пришло время шпаргалок Итак, впереди сессия, позади напряженная подготовка к ней родимой. Для дальнейшей подготовки к самому экзамену нам понадобятся шпаргалки, условно их можно разделить на «Бомбы» и «Учебники». Как вы будете волочить, и проносить учебники я не представляю, но некоторые умудряются, мы же поговорим о «Бомбах». «Бомба» это заранее

Тот же текст – но структурированный:

БЕСПРОБЛЕМНАЯ СДАЧА ЭКЗАМЕНОВ: ШПАРГАЛКИ ИЛИ МИКРОНАУШНИКИ ТЕХ?

БОТАНИКАМ - НЕ СЮДА!

Здесь никто не станет убеждать вас, что лучше никогда не писать шпаргалок и учиться, учиться и еще раз учиться. Если можете учиться, принося свою молодую жизнь на алтарь победы, знаете наизусть всю Большую Советскую энциклопедию, насмешливо глядите на преподавателя с выражением «Я знаю больше чем ты», тогда вы Анатолий Вассерман, зачем вы здесь? Из любопытства? Да ладно. Здесь мы поговорим о шпаргалках, и как правильно подготовиться к их использованию.

ПОДГОТОВКА ПОЧВЫ

Приучить преподавателя к вашим «странностям»

Для начала, каждый преподаватель должен знать, что вы человек чертовски рассеянный и параноидальный, вы постоянно роняете вещи, прячете от всех, что-то в одежде. Вы всегда закрываете рукой учебники и тетрадки, потому что вам кажется, что за вами подглядывают и списывают, делайте это достаточно часто, чтобы ваше поведение принималось как само собой разумевшееся, тогда пол дела сделано.

Создать электронный конспект

Параллельно с ведением конспекта старайтесь заносить всю информацию в персональный компьютер, если не сможете сами охватить весь объем, агитируйте других членов группы, мол,

Стандарты оформления текстовых документов

| Тип стандарта | Пример или описание | Комментарий |
|--|--|---|
| Государственные стандарты оформления текстовых документов | Межгосударственный стандарт ГОСТ 2.105-95 "Единая система конструкторской документации. Общие требования к текстовым документам". | Морально устарел. Не является обязательным в большинстве организаций. |
| Государственные стандарты оформления отдельных элементов текста | ГОСТ Р 7.0.5 - 2008 «Библиографическая ссылка. Общие требования и правила составления» | Актуален и требуется к исполнению в учебных заведениях. |
| Государственные и международные стандарты на отдельные виды документов | ISO/IEC 10303 Standard for the Exchange of Product Model Data (STEP) и его российский аналог ГОСТ Р ИСО 10303 «Системы автоматизации производства и их интеграция» | Актуален, особенно при работе с зарубежными фирмами. |
| Корпоративные стандарты | Чаще существуют в виде набора шаблонов с элементами фирменного стиля | Обычно требуется их жёсткое соблюдение для внешней документации |
| Правила оформления отдельных видов документов | Оригинал-макет учебного пособия. Правила разработки и оформления: учеб.-метод. пособие / В. Н. Козлов [и др.] – СПб.: Изд-во Политехн. ун-та, 2008. | Применимость и обязательность их определяется руководством. |
| Нормы оформления административных и юридических документов | Стандартные бланки документов, правила делопроизводства | Несоблюдение чревато лишними проблемами. |

Итоги:

- Текстовая информация – самый распространённый пока вид информации на компьютере.
- Текст хранится в памяти компьютера в виде последовательностей кодов символов. Наиболее распространённые коды – ASCII и Unicode.
- Программы воспринимают текст по-разному: текстовые редакторы – только с точки зрения содержания; текстовые процессоры – охватывают ещё и структуру текста, есть программы, выполняющие с текстом специфические операции (машинный перевод, к примеру).
- Восприятие текста во многом зависит от его структуры. Человек, планирующий зарабатывать на жизнь интеллектуальным трудом, должен уметь представлять текст в удобной для восприятия форме.
- Существуют разнообразные стандарты оформления текста. Их надо знать и соблюдать.