

Уровень представления данных

Кодировки, шифрование

Кодировки

- В 80-ых и 90-ых стандартом были 8-битные СИМВОЛЫ
- Большое количество кодировок, которые поддерживали различные алфавиты
- Специализированные кодировки, которые частично поддерживали некоторые наборы и были удобными

Кодировки

- Проблема заключается в том, что все алфавиты всех стран не помещаются в 8 бит
- Появились различные кодировки, притом для одного языка по несколько вариантов
- Cp866 – DOS
- Koi8-r – email & *nix
- Cp1251 – windows
- Mac

Кодировки

- Кодировки, которые при потере символа транслитерировались (koi8-r)
- Из-за большого количества кодировок возникала проблема их определения
- Из-за неправильного определения и автоматического перекодирования проблема усугублялась
- Большие сложности при использовании нескольких в одном документе



Unicode

- Разрабатывается некоммерческой организацией «Консорциум Юникода» (Unicode Consortium, Unicode Inc.)
- Стандарт кодирования, позволяющий представить символы практически всех письменных языков
- Коды разделены на несколько областей
- U+0000 до U+007F содержит символы набора ASCII

Unicode совершенствование стандарта

- 1.1 стандарт 1991—1995 годов.
- 2.0, 2.1 стандарт 1996 года.
- 3.0 стандарт 2000 года.
- 3.2, стандарт 2002 года.
- 4.0, стандарт 2003.
- 4.01, стандарт 2004.
- 4.1, стандарт 2005.
- 5.0, стандарт 2006.
- 5.1, стандарт 2008.
- 5.2, стандарт 2009.
- 6.0, стандарт 2010.
- 6.1, стандарт 2012.

Кодовое пространство

- разбито на 17 плоскостей по 2^{16} (65536) символов
- нулевая плоскость называется базовой, в ней расположены символы наиболее употребительных письменностей
- первая плоскость используется, в основном, для исторических письменностей
- вторая — для редко используемых иероглифов ККЯ
- третья зарезервирована для архаичных китайских иероглифов
- плоскости 15 и 16 выделены для частного употребления

<http://ru.wikipedia.org/wiki/%D0%AE%D0%BD%D0%B8%D0%BA%D0%BE%D0%B4>

Система кодирования

- буквы, содержащиеся хотя бы в одном из обслуживаемых алфавитов;
- цифры;
- знаки пунктуации;
- специальные знаки (математические, технические, идеограммы и пр.);
- разделители.

Модифицирующие символы

- Графические символы в Юникоде подразделяются на протяжённые и непротяжённые (бесширинные).
- Непротяжённые символы при отображении не занимают места в строке (например, знаки ударения)
- Протяжённые символы иначе называются базовыми (англ. base characters), а непротяжённые — модифицирующими (англ. combining characters); причём последние не могут встречаться самостоятельно.
- Например, символ «á» может быть представлен как последовательность базового символа «a» (U+0061) и модифицирующего символа «´» (U+0301) или как монолитный символ «á» (U+00C1).

Unicode

- Из-за неоднозначности написания при помощи модифицирующих символов применяется процесс нормализации
- Для двунаправленных текстов (например, Еврейского или Арабского), символы записываются так же, обратное отображение обеспечивает приложение

UTF-8

- Наибольшая совместимость со старыми ASCII системами
- Один символ занимает от 2 до 4 байт
- Размер определяется по первым битам

Unicode UTF-8:

0x00000000 — 0x0000007F: 0xxxxxxx

0x00000080 — 0x000007FF: 110xxxxx 10xxxxxx

0x00000800 — 0x0000FFFF: 1110xxxx 10xxxxxx 10xxxxxx

0x00010000 — 0x001FFFFFF: 11110xxx 10xxxxxx 10xxxxxx 10xxxxxx

Вопросы

- Зачем было создано несколько кодировок русского языка?
- Сколько бит занимает ASCII символ?
- Почему в Unicode существуют неоднозначности?
- Каким образом разделено кодовое пространство Unicode?

Упаковка

- Упаковка данных применяется также на представительском уровне
- Целью упаковки является повышение скорости передачи данных
- Это создаёт дополнительную нагрузку как на принимающую, так и на передающую сторону в зависимости от используемого алгоритма

Шифрование

- Данные проходят через большое количество узлов, каждый из которых может их сохранить/проанализировать и т.д.
- Возникает естественное желание исключить утечки информации
- Не все шифровальные алгоритмы и устройства разрешены законом

SSL

- Возможна инкапсуляция практически любого протокола: HTTP, FTP, POP3, IMAP
- Существует бесплатная реализация OpenSSL и программное обеспечение для создания зашифрованных туннелей для других протоколов

Вопросы

- Какие преобразования могут выполняться на представительском уровне?
- Чем упаковка отличается от шифрования?
- С какими целями шифруется весь трафик, а не только пароль?
- Каким образом можно прочитать нешифрованную информацию, передаваемую по сети интернет?