

## INTERNET-ТЕХНОЛОГИИ АДРЕСАЦИИ И ПЕРЕДАЧИ ДАННЫХ.

С.Е. Коваленко <sup>1</sup>, А.В. Дубель <sup>2</sup>, К.А. Забродская <sup>3</sup>

<sup>1</sup> – студент 4 курса, факультета ФБД, группы ДБУ-3, Белорусского государственного экономического университета

<sup>2</sup> – студент 4 курса, факультета ФБД, группы ДБУ-3, Белорусского государственного экономического университета

<sup>3</sup> – научный руководитель, ассистент кафедры информационных технологий, Белорусского государственного экономического университета, Минск, 220672, Партизанский пр., 26, тел. (8017) 249-19-81, e-mail: [zabrodskaya\\_k@bseu.minsk.by](mailto:zabrodskaya_k@bseu.minsk.by)

**Аннотация.** В работе рассматриваются способы организации передачи данных в Internet-пространстве. Рассказывается о системе адресации: URI, URC, URL, FTP, HTTP URLs, FTP URLs и др. Также идет речь об особенностях использования электронной почты: кодировка, почтовые программы и их настройка.

**Ключевые слова:** система адресации в Internet, методы доступа, сервер, клиент, протокол, доступ к удаленным системам, кодировка букв, рабочая станция, почтовая оболочка.

### 1. Введение.

В информационном WWW пространстве все ресурсы имеют единую адресацию, которая называется **URI** - Uniform Resource Identifiers. Адреса URI автоматически присваиваются всем ресурсам сети Интернет независимо от методов доступа к ним (HTTP, FTP, Internet mail и т.п.).

#### Терминология.

**URI.** *Uniform Resource Identifier.* Множество всех имен и адресов (names/addresses) ресурсов WEB пространства.

**URL.** *Uniform Resource Locator.* Множество URI схем которые дают точную инструкцию по доступу к ресурсу через Интернет.

**URN.** *Uniform Resource Name.*

1. Это некоторый URI, который дает расширенную инструкцию возможной передачи ресурса. (это относится и к URL). Например, **PURLs**.

2. Частичная схема передачи ресурсов, которая развивается в настоящий момент **IETF**. Позволяет использовать в интернетовских протоколах имена, которые являются нестандартными (например, содержат русские буквы).

**URC.** *Uniform Resource Citation* или *Uniform Resource Characteristics.* Множество пар метаданных attribute/value, которые описывают ресурс. Некоторые из них могут быть URIs различного вида. В эти писания могут быть включены, например, authorship, publisher, datatype, date, copyright status and shoe size (см. [Metadata and RDF](#)).

### 2. Адрес URL - Uniform Resource Locator

Система адресации и собственно сами адреса в HTML документах называются URL. Что такое URL? URL - это аббревиатура от *Uniform Resource Locator*. Адрес URL является сетевым расширением понятия полного имени файла в операционной системе (пути к файлу *filename*). В этой адресации кроме имени файла и директории, где он находится указывается сетевое имя машины, на которой этот файл расположен и метод доступа к файлу, который можно использовать для его просмотра или загрузки.

#### 2.1. Методы доступа.

С помощью URL, описывающих различные методы доступа к удаленным системам, можно запускать удаленные программы и передавать им значения. На этом принципе построены шлюзы в другие интернетовские сервисы, например, *finger*, *archie*,..... и др.

Рассмотрим несколько наиболее часто используемых типов URL адресов и методов доступа.

#### 2.2. Файлы.

Допустим, файл с именем "internet.zip" лежит на FTP-сервере ftp.ict.nsc.ru в директории /pub/winsite/www/. Тогда URL адрес этого файла будет выглядеть так:

```
file://ftp.ict.nsc.ru/pub/winsite/www/internet.zip
```

Пример URL адреса директории, в которой лежит файл:

```
file://ftp.ict.nsc.ru/pub/winsite/www/
```

Понятно, что URL адрес корневой директории FTP сервера ftp.ict.nsc.ru выглядит вот так:

```
file://ftp.ict.nsc.ru/
```

Кстати вместо сетевого имени машины ftp.ict.nsc.ru можно указать ее IP-адрес 193.124.243.76. Это часто бывает полезно для ускорения работы с удаленными системами при невысокой скорости передачи данных, поскольку в этом случае нет необходимости обращаться к службе DNS для поиска сетевого адреса, по которому идет маршрутизация IP-пакетов.

### 2.3. FTP адрес.

Если вам необходимо явно указать, что выбранный вами файл следует загрузить с FTP сервера на машину клиента, то в URL адресе файла следует явно указать тип протокола, например:

```
ftp://ftp.ict.nsc.ru/pub/winsite/www/inter.net.zip
```

### 2.4. Gopher URLs

Адреса Gopher URL's не так разнообразны, как файловые. Это связано с ограниченностью этого сервиса. Для того, чтобы задать, например, gopher сервер Новосибирского государственного университета gopher.nsu.ru, необходим URL:

```
gopher://gopher.nsu.ru/
```

Некоторые gopher сервера могут находиться на нестандартном номере порта (по умолчанию обычно используется 70 порт). В этом случае он должен явно указываться:

```
gopher://gopher.banza.com:7070/
```

Здесь 7070 - номер порта.

### 2.5. HTTP URLs

HTTP -- HyperText Transport Protocol (протокол передачи гипертекста). HTTP сервера обычно используются для предоставления гипертекстовых документов. Такие документы, в отличие от обычных, имеют ссылки на другие документы (не обязательно расположенных на этом же сервере) и состоят из текста, графики, звуков, анимации.

Для вызова документа "url.html", который находится в директории /win/docs/html/ WWW сервера www.ict.nsc.ru необходим следующий URL адрес:

```
http://www.ict.nsc.ru/win/docs/html/url.html
```

По умолчанию все программы по протоколу HTTP ищут WWW сервер на 80 порту, но можно и явно указать порт (также как и в gopher'e).

```
http://www.weblab.com:1234/pub/files/foobar.html
```

Частичные URL используются только при протоколом передачи гипертекста (HTTP). Если вы внимательно посмотрите на исходники какого-нибудь гипертекстового документа и обратите внимание на то, как указаны ссылки на другие URL, то заметите, что встречаются два вида:

```
<A  
Href="http://www.w3c.net/WWW/url.html">URL</A>  
<A  
Href="docs.html">Документация</A>
```

Первый пример - это полный URL, а второй - частичный. Частичный URL указывает на документ, который находится на том же сервере и в той же директории, где и документ, в котором встречается эта ссылка. Так, например, если документ, где вы нашли эти две строчки, имел URL http://www.w3c.net/WWW/foo.html, то полный URL у второго частичного должен выглядеть как http://www.w3c.net/WWW/docs.html.

Сокращенные URL адреса определяются по правилам, принятым в операционной системе UNIX. Например, адрес ../win/file.html означает переход на одну директорию вверх и выбор файла из поддиректории win с именем file.html или адрес /docs/ball.gif - переход в корневую директорию и выбор файла с полным именем /docs/ball.gif.

### 2.6. Другие URL адреса

Приведем несколько других типов задание доступа к удаленным системам:

```
<A HREF="mailto:...">
```

Задание адреса электронной почты для автоматической отправки сообщения через просмотрщик:

```
<A HREF="mailto:root@ict.nsc.ru">  
root@ict.nsc.ru</a>
```

```
<A HREF="news:...">
```

Задание связи с сервером Newsgroup

```
<A HREF="telnet://...">
```

Задание связи через TELNET.

```
<A HREF="wais://...">
```

Задание связи с сервером WAIS.

## 3. Электронная почта (E-mail).

Почта является самой популярной и распространенной службой **Internet** в нашей стране. К ней все привыкли и уже не могут без нее жить.

Родоначальником массового распространения **E-mail** в нашей стране является сеть EuNet/Relcom, которая всем диктовала свои правила использования. Это было нормально и все с этим мирились пока в стране сеть **Relcom** являлась единственной сетью, предоставляющей телекоммуникационные услуги, и операционная система **MS DOS** являлась основной системой на которой работали пользователи.

На самом деле электронная почта как средство передачи сообщений между пользователями ЭВМ, в том числе и у нас в стране, возникла задолго до появления **Internet** и **Relcom**'а.

Если Вы работаете только с английским алфавитом и отправляете свои сообщения тоже, используя только английский шрифт, то у Вас вряд ли возникнут проблемы с использованием и установкой на свой компьютер программы поддержки **E-mail**. Основные проблемы возникнут, когда Вы захотите отправить сообщение, используя кириллический (русский) шрифт. Все дело в том, что у нас в стране не существует единого стандарта кодировки русских букв.

### 3.1. Русские кодировки

В настоящее время в нашей стране на разных компьютерах используются множество различных кодировок русских букв: самые популярные из них это

- Кодировка KOI8 (RFC1489) - используется в большинстве **OS UNIX** (charset=KOI8-R);
- Кодировка MS DOS (CP866) - используется в **OS MS DOS 6.22** для **IBM PC** (charset=x-CP866);
- Кодировка MS Windows (CP1251) - используется в **MS Windows** (charset=Windows-1251);
- Кодировка ISO-8859-5 - используется в **DEC** (charset=ISO 8859-5);

Кроме них существуют еще кодировка **MacOS**, а также почти забытая альтернативная кодировка (**CP855**), и уже совсем забытые кодировки основная (или болгарская) и **ГОСТ**.

**Замечание:** в настоящий момент в сети **Internet** для этой кодовой таблицы используются и другие синонимы (alias) для указателя charset.

Из всех перечисленных кодировок только кодировка **ISO8859-5** "узаконена" Международной Организацией по Стандартизации (ISO), но к сожалению она очень редко используется на современных компьютерах.

Сеть **Internet** развивалась первоначально как сеть **UNIX** машин и унаследовала все свойства представления информации в **OS UNIX**. Для передачи информации первоначально использовался только 7 битный код (т.е. передавалась только первая половина кодовой таблицы). Позже были разработаны различные способы передачи всей кодовой таблицы (transfer encoding=8 bit).

Наследие **UNIX**'а коснулось и системы передачи электронной почты. Неписанным телекоммуникационным стандартом передачи русских букв изначально стал стандарт **OS UNIX** кодировка **KOI8**. Практически все почтовые оболочки (изначально разработанные для **MS DOS**) перекодировали **DOS**'овскую кодировку **CP855** в **KOI8** и после этого передавали сообщение.

Такое единообразие нарушилось с появлением в нашей стране других сетей, входящих в систему **Internet**, таких как **HEPNet (Radio-MSU)** (этой сети принадлежит "Сеть Internet Новосибирского научного центра"), **FreeNet**, **RUNNet** (сеть университетов России) и др., и новых операционных систем, таких как **MS Windows NT**, **MacOS**, **OS/2** и др. и системы передачи всей кодовой таблицы. Начиная с 1995 года старая система передачи русских сообщений стала постоянно давать сбои и корреспонденты то и дело вместо русского текста стали получать абракадабру, которую правда можно было расшифровать, или хуже всего сплошные квадратики (или точки), которые уже не поддавались расшифровке. Тем самым единый способ перекодировки русских букв оказался неприемлемым для использования в передаче почтовых сообщений.

Для правильной работы, используемая Вами оболочка для передачи почтовых сообщений, которые содержат русские символы, должна правильно выставлять указатель кодовой таблицы (**Code Page**) (transfer encoding=8 bit Charset="УКАЗАТЕЛЬ ТАБЛИЦЫ") (см. список синонимов - alias - кодовых таблиц).

К сожалению старые **DOS**'овские почтовые системы не понимают различных

систем кодировок. Поэтому, если Ваш корреспондент сидит за рабочей станцией, которая является DOS'овской машиной, то, чтоб не создавать ему лишних трудностей, рекомендуется отправлять почту в кодировке **KO18-R**, а DOS'овская почтовая оболочка аккуратно переведет ему Ваше сообщение в альтернативную (или в кодировку CP866) кодировку.

### UNIX Mail = Internet Mail

Для того, чтоб пользоваться электронной почтой в сети **Internet** Вы должны быть зарегистрированным пользователем (user'ом) на какой-нибудь ЭВМ, работающей в операционной системе UNIX (такая ЭВМ в сети **Internet** называется host'ом). Система передачи почтовых сообщений операционной системы UNIX по-прежнему является неписанным стандартом сети **Internet**.

Эта ЭВМ может находиться в любой точке земного шара, но должна быть подключена к сети **Internet**, а Ваша рабочая станция должна иметь с ней IP-соединение (IP - Internet Protocol).

Для того, чтоб стать зарегистрированным пользователем Вам нужно обратиться с просьбой к системному администратору ближайшей от Вас (ехать для этого в Америку по-моему неразумно) UNIX машины, о том, что бы он завел Вам личный account (user name, которое будет Ваши адресом электронной почты и password - пароль). После этого Вы становитесь полноправным пользователем электронной почты сети **Internet**.

На самом деле Ваш почтовый адрес (электронный адрес) составляется из Вашего user-name и IP адреса UNIX машины, отвечающей за прием Вашей почты, например, User@host.domen.nsc.ru

Таким образом, число Ваших почтовых адресов (почтовых ящиков) равно числу ЭВМ, на которых Вы являетесь пользователем. Что бы избежать путаницы в своих почтовых ящиках, желательно пользоваться только одним, а по всем другим поставить автоматическую пересылку в выбранный Вами почтовый ящик и проверять почту только в нем.

Чтобы отправить или получить почтовое сообщение Вам нужно сесть за свою рабочую станцию (или за компьютер своего знакомого), которая имеет выход в **Internet** и с помощью программы telnet соединится со своим host'ом. После этого

набрать команду mail и начать работать с почтовой службой **Internet**.

Конечно, работать с помощью командной строки OS UNIX крайне неудобно, да и уже непривычно после многолетней работы на PC.

Для удобства пользователей разработана масса различных почтовых оболочек, работающих под всеми существующими операционными системами.

В OS UNIX для алфавитно-цифровых мониторов типа vt100 имеется почтовая оболочка pine, которая работает аналогично системе **Pmail**

К сожалению фирма Microsoft (автор MS DOS) объявила о прекращении каких-либо новых разработок для MS DOS и о постепенном свертывании разработок для MS Windows 3.11, поэтому работая с этими OS уже навряд ли можно будет использовать новые возможности, появляющиеся в сети **Internet**.

Естественно любая программная оболочка, особенно если она установлена не на UNIX-машине, нуждается в предварительной настройке (см. раздел "Настройки почты"). Если Вы работаете с терминалом UNIX-машины, то настраивать почтовую систему, как правило не нужно, поскольку UNIX-машина самостоятельно умеет принимать и отправлять почту.

### 3.2. Netscape Navigator

Если на Вашей рабочей станции установлен Netscape Navigator, то выход на электронную почту Вам обеспечен почтовой оболочкой, входящей в его состав. Netscape Navigator версии 3.0 автоматически распознает кодировки **KO18**, **Win1251** и **ISO8859-5**. Правда чтобы прочитать письмо написанное в этих кодировках на Вашей рабочей станции должны быть установлены соответствующие шрифты, которые в свою очередь должны быть подключены к Netscape Navigator (меню: Options, General, Preferences, Fonts). Шрифты **KO18** можно взять с FTP сервера.

Для более корректной работы с русским языком следует использовать версию Netscape Navigator 3.0 или 4.01 (beta), которую можно загрузить с FTP сервера ИВТ СО РАН.

Имея в своем распоряжении шрифты **KO18**, Вы можете читать практически любую русскоязычную почту (кроме той, что пришла, например, в альтернативной кодировке или в кодировке ISO 8859-5). Отметим, что для версии 4.0 и выше при

работе под MS Windows 95/NT нет необходимости загружать шрифты KOI8 на Ваш компьютер, так сам просмотрщик делает необходимую перекодировку.

Подготовку почтовых сообщений для Netscape Navigator под MS Windows 3.11 можно организовать, используя драйверы клавиатуры KOI8, выбор которых зависит от вида русификатора, установленного на вашей машине:

- [KOI8-R keyboard for ParaWin](#)
- [KOI8-R keyboard for CyrWin 4.0](#)
- [KOI8-R keyboard variant for WinKey](#)
- [KOI8-R keyboard for Win95](#)

Гораздо сложнее с отправкой почты под MS Windows 95 с Netscape Navigator 3.0, так как он категорически не желает воспринимать драйвер клавиатуры для кодировки KOI8 и Вы вынуждены набирать свои сообщения только в кодировке MS Windows. К сожалению такое почтовое отправление правильно поймет почтовая программа, которая знает эту кодировку. Большинство почтовых программ, работающих под DOS'ом к сожалению этого не понимают и Ваш корреспондент будет всегда стоять перед проблемой перекодировки.

Если Вы получили сообщение в кодировке KOI8, а Вам для работы нужен текст в альтернативной кодировке, то для перекодировки можно использовать программу [cat.exe](#).

Обойти эту проблему можно, установив на своей машине [Local proxy 1251 -- koi8r for Win95](#), осуществляющий перекодировку почтовых сообщений (протоколы: [SMTP](#) - отправка, [POP3](#) - получение почты).

Для начинающих рекомендую посмотреть статью Егора Чемисова о настройках и работе с Netscape Navigator 3.0 на сервере <http://www.binet.lv/>: кодировка [MS Windows] и [KOI8]

### 3.3. MS Internet Mail

Если на Вашей рабочей станции установлена OS Windows 95 (или NT), то разумнее всего использовать оболочку MS Internet Mail, которая является компонентой Microsoft Internet Explorer полной версии 3.0, который можно получить с [WWW сервера фирмы Microsoft](#).

Для MS Windows 95/NT Microsoft Internet Explorer доступен также на [FTP сервере ИВТ СО РАН](#).

Microsoft Internet Explorer использует два режима передачи расширенной таблицы: [MIME](#) и [UUENCODE](#).

Если Вы укажете ему использовать MIME, правильно работает с русскими текстами (выставляет нужный charset), однако для прицепленных файлов он использует методы передачи бинарных файлов по электронной почте Quoted Printable и base64, о которых мало кто в нашей стране знает.

В режиме UUENCODE MS Internet Mail использует 7 битную передачу с кодировкой основного текста, не используя charset, и передает текст как он есть. На этом могут споткнуться "умные" DOS'овские почтовые программы, которые захотят перекодировать Ваше сообщение в альтернативную кодировку. Поэтому в этом случае, чтоб быть правильно понятым следует использовать режим передачи почты в кодировке KOI8.

Отметим, что для использования MS Internet Mail нет необходимости загружать фонты и драйвер клавиатуры KOI8.

Русскую инструкцию по использованию MS Internet Mail можно найти на сервере <http://www.binet.lv/>: кодировки: [MS Windows] и [KOI8], ее автор: Егор Чемисов (<mailto:egor@binet.lv>).

Кстати, если Вам понадобится послать или прочитать сообщение в кодировке ISO8859-5, то можно воспользоваться [MS Exchange Translation Tables for koi8-r and iso8859-5](#).

Существует так же вариант программы MS Internet Mail в составе Microsoft Internet Explorer для MS Windows 3.11 на WWW сервере фирмы Microsoft версии 2.01.

Вместе с почтовой программой и browser'ом пакет содержит поддержку протоколов TCP/IP, SMTP и DIAL-UP соединения. Вместе с русской версией пакета поставляется файл автоматической перекодировки в KOI8.

### 3.4. Eudora MAIL

Почтовая оболочка Eudora является небольшим свободно распространяемым программным продуктом (Free Ware), работающим под MS Windows 3.11 и MS Windows 95 (существует также принципиальная возможность запустить эту оболочку и под MS DOS, установив при этом пакет Winsoket - резидентная программа занимающая около 100 Kb оперативной памяти).

Основное преимущество оболочки Eudora - то, что она является многопользовательской

Eudora получает и отправляет почту в том виде в котором она подготовлена (в этом

отношении правильно работает с русскими шрифтами), самостоятельно определяет, используется ли в отправляемом сообщении вторая половина таблицы (русские шрифты) и выставляет соответствующий charset.

К сожалению Eudora умеет работать только с одной расширенной кодовой таблицей (в стандартном варианте это кодировка ISO8859-2). Программу Eudora, настроенную на передачу кодировки **KO18-R** можно взять с FTP сервера Института.

Кроме того, что Eudora небольшая программа (сама программа имеет размер 700 Kb, Help File - 1 Mb и около 400 Kb английский спеллинг, который можно удалить после установки), основным ее достоинством является то, что она может работать в многопользовательском режиме (в отличие от Netscape Navigator или MS Internet Mail).

Для установки программы Eudora нужно распаковать ее в директорию, установить драйвер клавиатуры **KO18** и хотя бы один из шрифтов **KO18** (а лучше два: один для просмотра, другой для печати), которые можно взять с FTP сервера.

Пользовательский режим определяется командной строкой запуска программы:

```
c:\eudora\eudora.exe c:\usermail
c:\userini\usereudora.ini
```

где

- c:\eudora - каталог с программой Eudora;
- c:\usermail - почтовый (рабочий) каталог пользователя User;
- c:\userini\usereudora.ini - файл настроек пользователя User.

Данная командная строка записывается в соответствующий PIF файл для MS Windows 3.11 и в параметры ярлыка для MS Windows 95.

Создать первоначальные копии файлов настроек пользователя можно, размножив файл настроек eudora.ini из каталога, где находится Eudora.

### 3.5. Другие почтовые программы

Существующие версии программы WinPmail (Pegasus Mail for Windows), в отличие от DOS'овского варианта (который использует стандартную перекодировку KO18 -- CP855) плохо работает в стандартном

режиме с русскими буквами. Использование способа передачи **MIME**, как уже отмечалось не уживается со старым программным обеспечением.

Среди других почтовых программ можно предложить оболочку **Mail Connection**. Русское описание этой программы.

Кстати, распространенная для UNIX систем почтовая программа **PINE** (свободно распространяемая оболочка, разработанная в Вашингтонском университете) работает так же и под MS Windows.

### 3.6. Настройка программ.

Любая почтовая оболочка, если она работает не на UNIX-машине, нуждается в настройке. Первое и самое главное условие, которое выдвигает Ineternet, - это то, что на вашей машине должен быть установлен межсетевой протокол **TCP/IP**. Совершенно не важно на каком физическом уровне соединения он поддерживается (через локальную сеть организации, по телефонному каналу и пр.) - важно только то, что это физическое соединение должно доходить до ЭВМ, которая называется **Name Server**.

Name Server - это ЭВМ, в память которой занесено имя Вашей рабочей станции и ее IP-адрес.

В настройках почтовой программы **В** должны указать:

- Ваш почтовый адрес или account, например, user@net.ict.nsc.ru - email address user - account.
- Имя или IP-адрес UNIX-машины, в которой находится Ваш почтовый ящик, например, net.ict.nsc.ru - имя 193.124.243.78 - IP-address.
- Имя или IP-адрес SMTP сервера (SMTP server), UNIX-машины, которая будет отправлять Вашу почту, например, mailhost.ict.nsc.ru - имя 193.124.243.33 - IP-address.

### ЛИТЕРАТУРА

<http://www.ict.nsc.ru/win/docs/html-gd/url.html>  
<http://www.ict.nsc.ru/rus/inter/e-mail.html>