Lotus Notes R5 on Mac

История о том, как я пытаюсь привинтить Lotus Notes R5 на Мак

Эту страничку я создал с целью решения проблемы несовместимости Lotus Notes Client for Mac and PC. Может вы найдете для себя что-то новое.

Часть 1

Есть Lotus Notes R5 Server Russian.

Очень хочется подлючить к этому серверу и маковских пользователей.

Фирма Lotus в былые времена столкнувшись с проблемой реализации много-язычных документов придумала для кодировки текста формат LMBCS (Lotus Multi-Byte Character Set). Основная идея LMBCS состоит в том, что буква записывается при помощи двух или трех байтов, в зависимости от языка.

Вот структура формата:

LMBCS data format

Each LMBCS character has the basic structure of:

[O] [D1] [D2]

where

[O] = optional optimization byte, in the range 0x00-0x1F (excluding 0x09=H, 0x0A=LF and 0x0D=CR). The optimization byte may be defaulted to a specified value (**in Notes/Domino this is 0x01**).

[D1] = Data byte, which can be any character in the range 0x01-0xFF. When the data byte is not preceded by an optimization byte, it can have the value 0x09, 0x0A, 0x0D or any ASCII character in the range 0x20 - 0x7F inclusive. It can also have a value in the range 0x80-0xFF <>

[D2] = Second data byte for use in DBCS encodings only and therefore optional

The optimization groups indicate the follwing encodings (note that this applies ONLY to data in the range 0x80-0xFF):
0x01 - Codepage 850, ICU alias "ibm-850"
0x02 - Codepage 851, currently not implemented in ICU
0x03 Codepage 1255 (pre-Euro), ICU alias "ibm-1255"
0x04 - Codepage 1256 (pre-Euro), ICU alias "ibm-1256"
0x05 - Codepage 1251 (pre-Euro), ICU alias "ibm-1251"
0x06 - Codepage 852, ICU alias "ibm-852"
0x07 - UNUSED
0x08 - Codeage 1254 (pre-Euro), ICU alias "ibm-1254"
0x09 - UNUSED (HT)
0x0A - UNUSED (LF)
0x0B - Codepage 874 (pre-Euro), ICU alias "ibm-874"
0x0C - UNUSED
0x0D - UNUSED (CR)
0x0E - UNUSED
0x0F - UNUSED
0x10 - Shift-JIS , ICU alias "ibm-943"
0x11 - KS , ICU alias "ibm-949"

PDF created with pdfFactory Pro trial version www.pdffactory.com

0x12 - Big-5 , ICU alias "ibm-950"
0x13 - GBK , ICU alias "ibm-1386"
0x14 - Unicode compatibility region (Unicode encoded as big-endian UTF-16)
0x15 - 0x1F - UNUSED

Ну прямо UNICODE :-) Действительно похоже, но внедрили в жизнь раньше.

К тому же и сама фирма Лотус говорит о том, что маковский клиент очень хорош собой и великолепно понимает все языки. Наверное это правда.

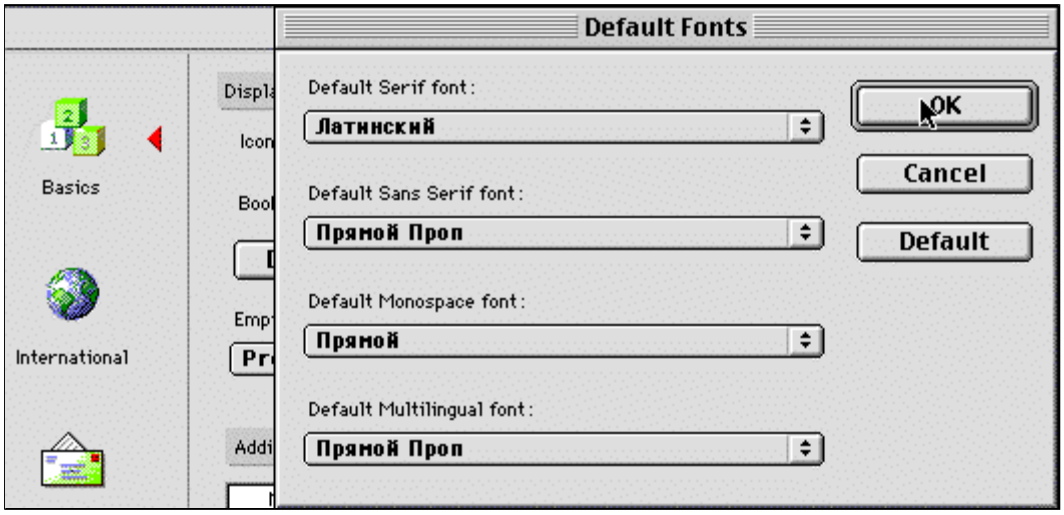
А вот что я имею на практике:
Чтение данных из реальной базы.

На маке стоит MacOS 9.04, **Primary system script - Cyrillic (script 7)**. Как и требует того фирма Apple.
То, какая основная языковая среда установлена в системе, имеет большое значение !!!

После инсталяции клиент правильно понимает свой location и правильно прописывает это в настройках

File => Preferences => User Preferences ... => International

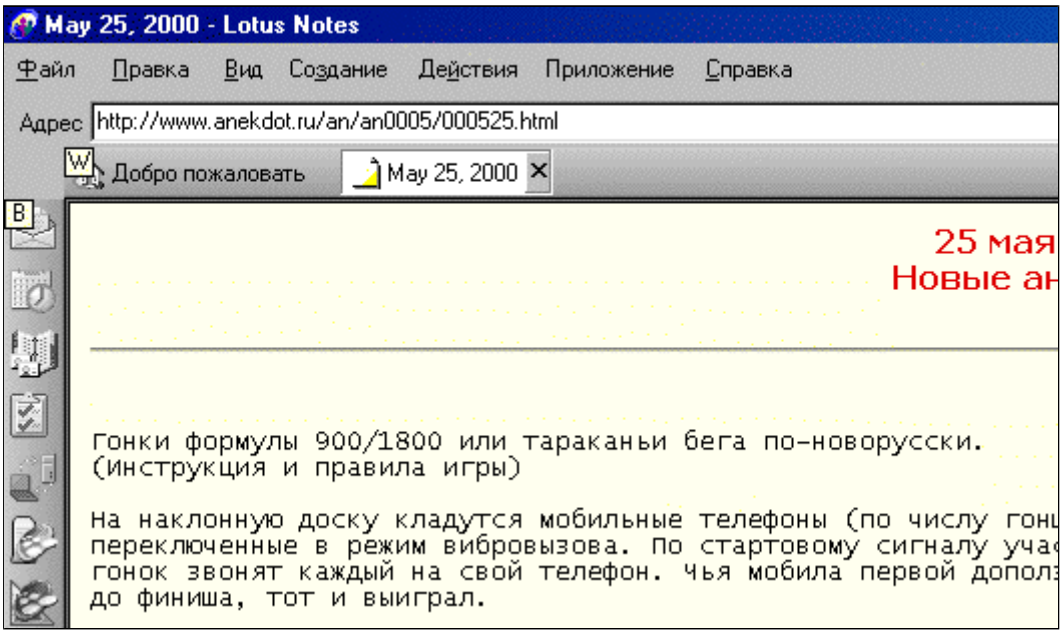
Также я прописываю кириллические шрифты. Сам он до этого не додумывается. Этого правда от него никто и не требует. :-)



File => Preferences => User Preferences ... => Basics => Default fonts ...

Все мы знаем, что в пятой версии есть web-броузер. И это просто прекрасно. Вот как он работает в версии для PC.

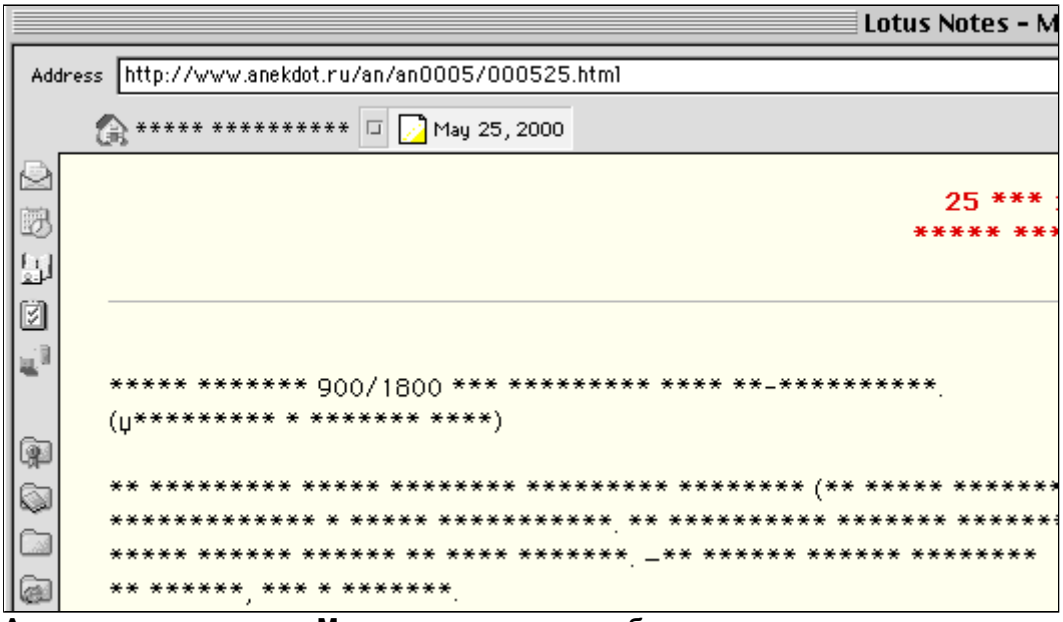
PDF created with pdfFactory Pro trial version www.pdffactory.com



Вот что мы видим на PC при просмотре веб-страниц

и нам это очень нравится

А теперь мы с вами хотим то же самое увидеть и на маковском клиенте. Пробуем ...
... и получаем вот что:

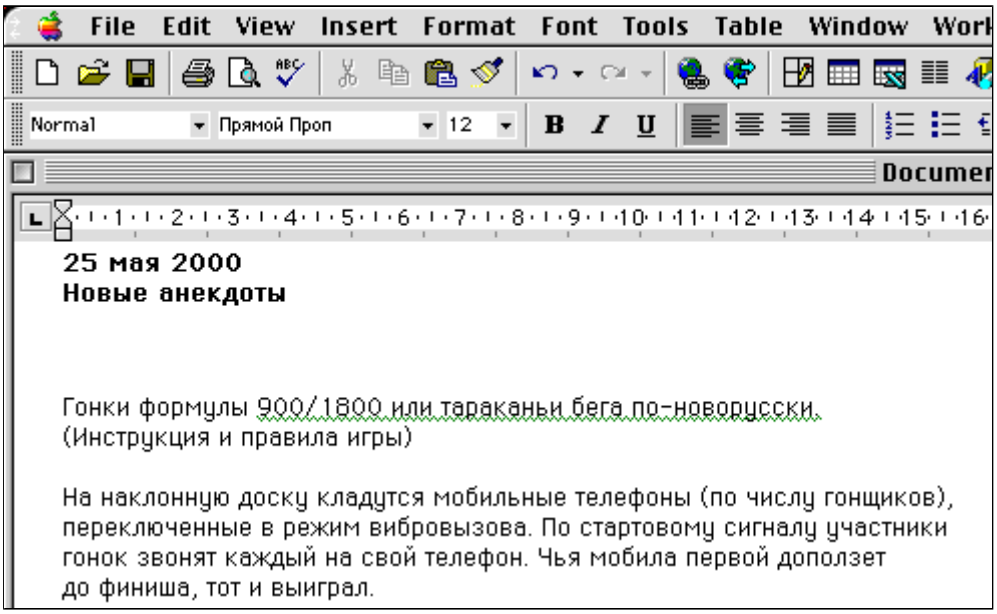


А вот что мы видим на Маке при просмотре веб-страниц

Видны только латинские символы !!! (0x20?0x1F плюс 0x09=HT, 0x0A=LF и 0x0D=CR).

Копирую полученный текст через клипборд в **Microsoft Word 98 for Mac** и о чудо:

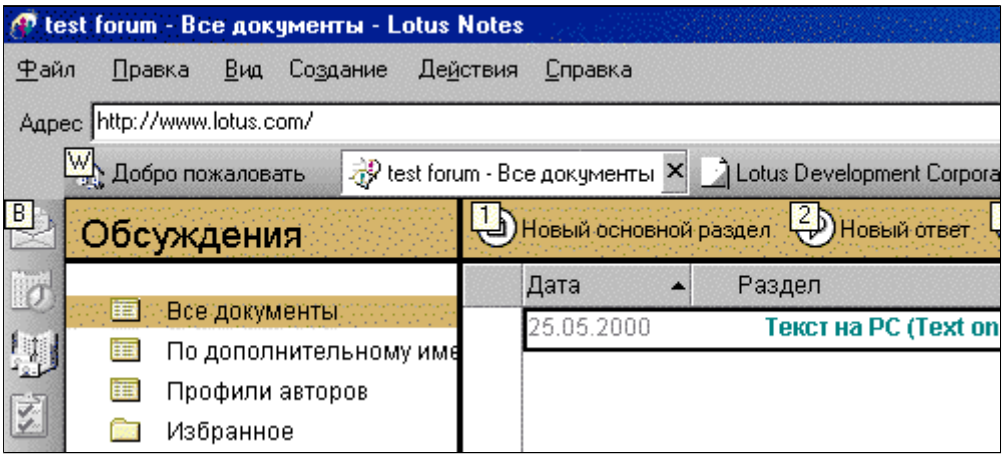
PDF created with pdfFactory Pro trial version www.pdffactory.com



MS Word с энтузиазмом сделал из звездочек удобочитаемый текст :-)

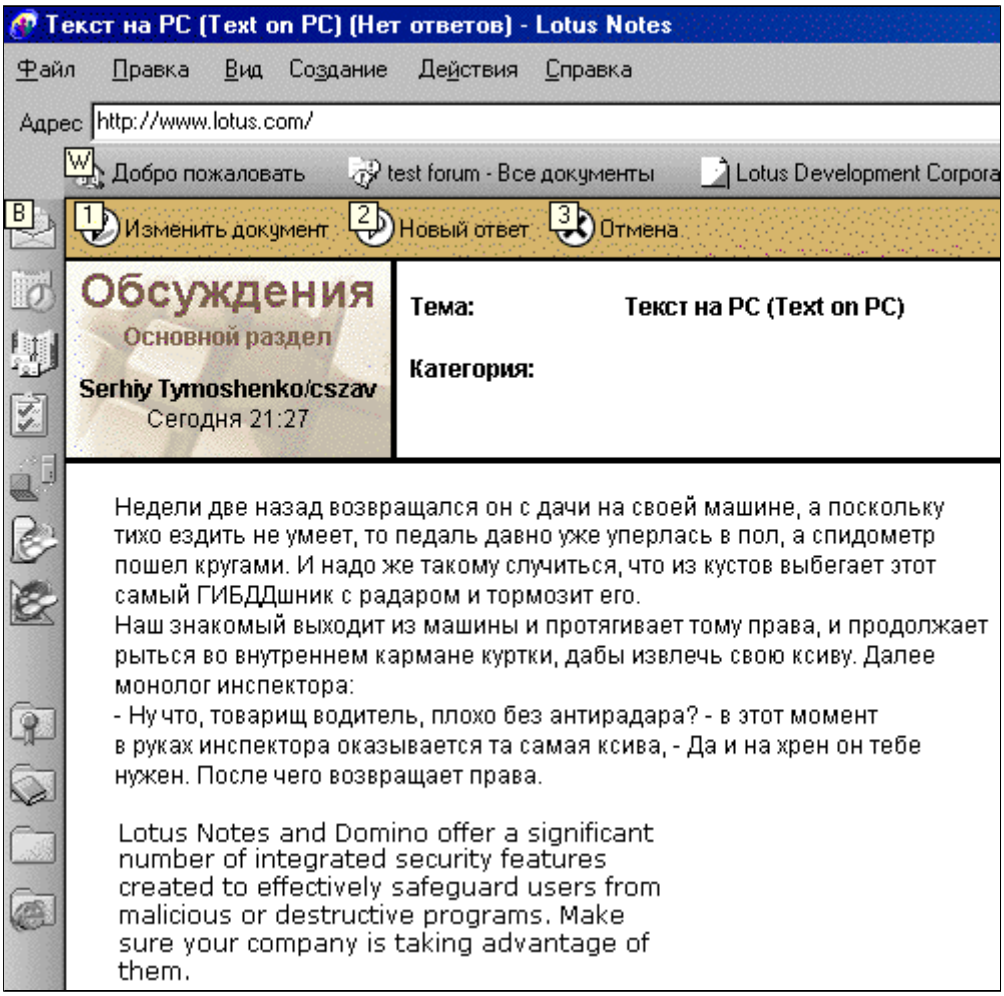
Важно отметить, что такой фокус проходит **только с Вордом.** Ни SimpleText,ни BBEdit на такое не способны.
Они показываю вместо кириллицы те же звездочки. Я думаю, что успех Ворда в его "дружбе" с технологией
UNICODE и умением корректно понимать и конвертировать содержимое клипборда. Вот вам и глючный
Билли. :-) При этом маковский клиент правильно понимает в какой именно кодировке он получает веб-страничку
и все правильно конвертирует. Я проверял сей факт вышеописаным "методом клипборда". Я решил проверить
работу веб-броузера с различными кодировками, а именно его способность эти кодировки распознавать.
Все замечательно работает как в родной маковской кодировке MacCyrillic (port 8082), так и в CP1251
(port 8080), и в KOI-8 (port 8083). Все распознается, можно заменить кодировку и страничка правильно перекодируется. Вот только на экране видны одни зведочки. При этом лотусовский клиент не почему-то не учитывает такой важный параметр, как "**Import/Export Characters: Codepage 1251"**, который я указал
выше.

Вы мне скажете, что это явный глюк встроеного броузера. Хорошо. Приведу пример работы с реальной базой. Быстренько клепаю из стандартного шаблона базу форума.

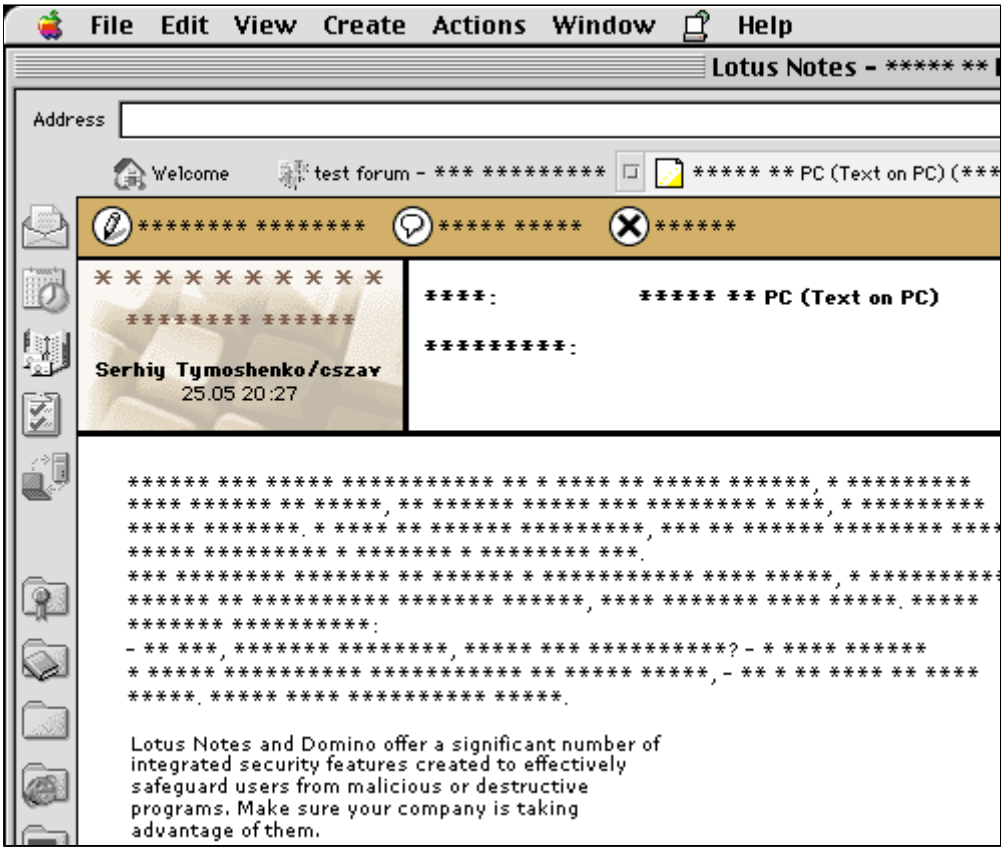


Создаю в этой базе запись с использованием кириллических и латинских букв...

PDF created with pdfFactory Pro trial version www.pdffactory.com



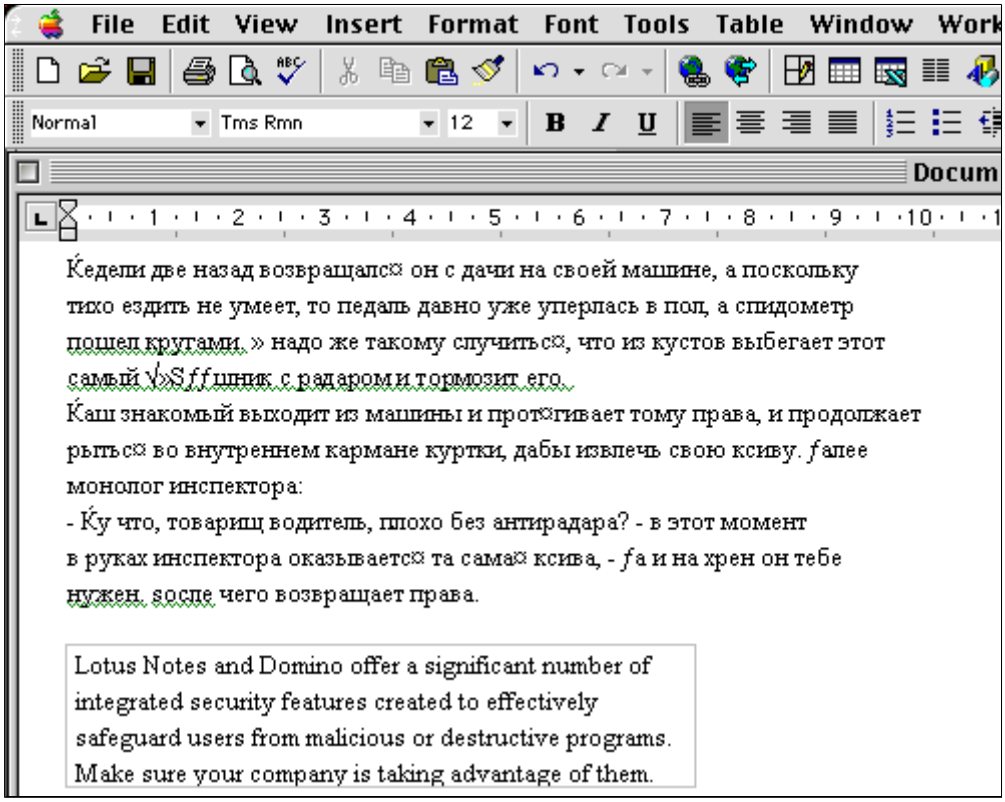
И пробую прочитать эту базу на маке ...



PDF created with pdfFactory Pro trial version www.pdffactory.com

Все те же звездочки.

А теперь я копирую "звездный" текст в клипборд и смотрю на него в Ворде.



С Вордом все понятно. Это кодировка Win-1251.

Кодировка закономерна. Она указана мною в национальных настройках для экспорта/импорта информации. Забавно, что эта настройка действует только на информацию в лотусовских базах и не работает с www, новостями и почтой.

Если же я попробую указать в настройках экспорта/импорта "Mac Script 7 (Cyrillic)", то вообще ничего не смогу прочитать. В клипборд будет скопирована полная белиберда. Хотя по книжкам нужно делать именно так и никак иначе. Имеем явное расхождение теории с практикой.

Часть 2 . **Форум**

PDF created with pdfFactory Pro trial version www.pdffactory.com