Лабораторная работа N1
Взлом моноалфавитного подстановочного шифра методом частотной атаки

Цель работы: ознакомиться на практике с использованием частотной криптоатаки при взломе подстановочных шифров.

Исходные данные:
Зашифрованный текст, перечень наиболее часто встречающихся букв в тексте, перечень наиболее часто используемых в русском языке букв.

Выходные данные:
Расшифрованный текст.

Теоретические основы:

Моноалфавитный подстановочный шифр - шифр, в котором каждой букве исходного алфавита поставлена в соответствие одна буква шифра.

Например, возьмем слово «КУКУРУЗА». Пусть букве «К» текста соответствует буква «А» шифра, букве «У» текста соответствует буква «Б» шифра, букве «Р» текста соответствует буква «В» шифра, букве «З» текста соответствует буква «Г» шифра, букве «А» текста соответствует буква «Д» шифра. После подстановки букв шифра вместо букв исходного теста слово «КУКУРУЗА» в зашифрованном виде будет выглядеть как «АБАБВБГД».

Недостатком подобного шифрования является то, что, если какая-то буква встречается в исходном тексте чаще всего (например, буква «О» в русском алфавите), то и соответствующая ей буква шифра в зашифрованном тексте также встречается чаще всего.

В нижеприведенной таблице приведены частоты встречаемости букв в английском тексте (в процентих):

Высокая		Средняя		Низкая	
E	12,31	L	4,03	B	1,62
T	9,59	D	3,65	G	1,61
A	8,05	C	3,20	V	0,93
O	7,94	U	3,10	K	0,52
N	7,19	P	2,29	Q	0,20
I	7,18	F	2,28	X	0,20
S	6,59	H	2,25	J	0,10
R	6,03	W	2,03	Z	0,09
H	5,14	Y	1,88		

Зная частоты наиболее встречающихся букв и подсчитав, какие буквы чаще всего встречаются в шифровке, криптоаналитик может подобрать расшифровку для некоторых букв текста. Затем, анализируя короткие слова, найти еще буквы, истинные значения которых можно с высокой степенью уверенности предугадать. Например, если уже расшифрована буква «О» и в тексте есть слово «ОЫО» (подчеркнуты уже расшифрованные буквы), то, скорее всего, шифру «Ы» соответствует буква «Н» в исходном тексте («ОНО»). Чем дальше расшифровывается текст, тем легче идет процесс расшифровки.

Методические указания:

1. Запустить на выполнение файл labw01.exe
   На экране появится окно выполнения лабораторной работы (рис. 1):

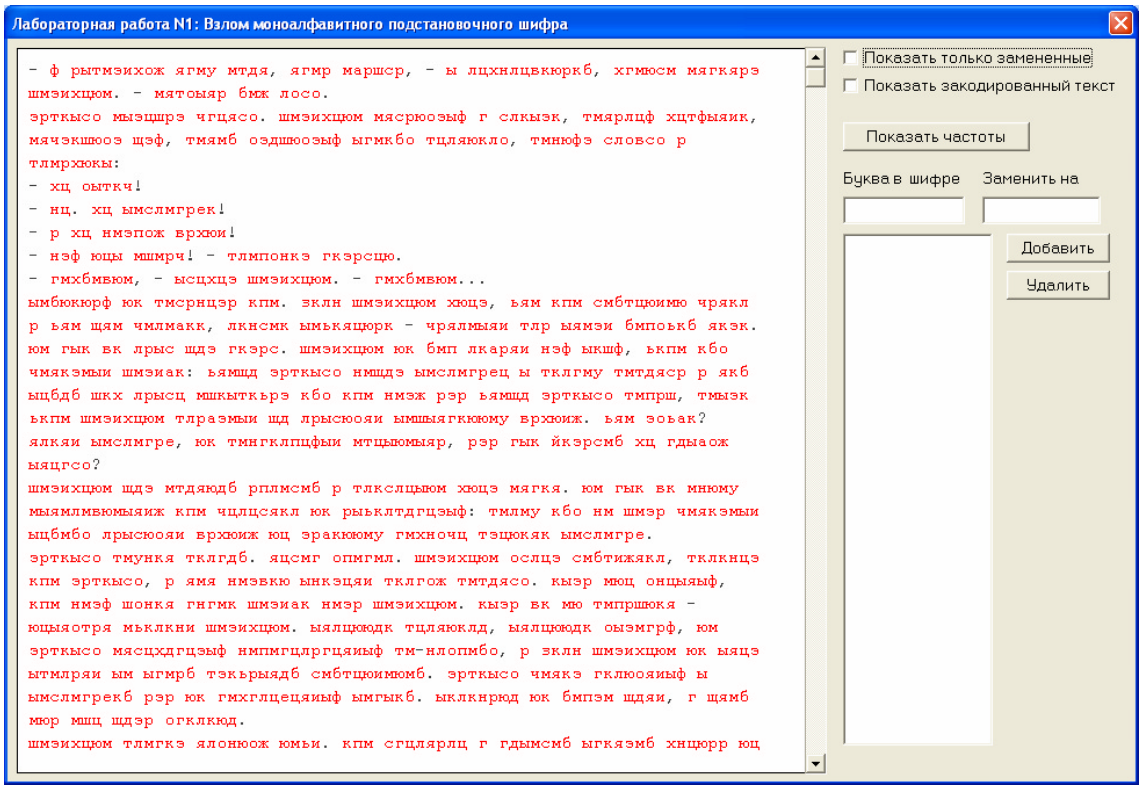

Рисунок 1. Окно выполнения лабораторной работы

В левой части окна находится зашифрованный текст (буквы, выделенные красным цветом). В процессе расшифровки расшифрованные (правильно или неправильно) буквы текста меняют цвет с красного на черный.

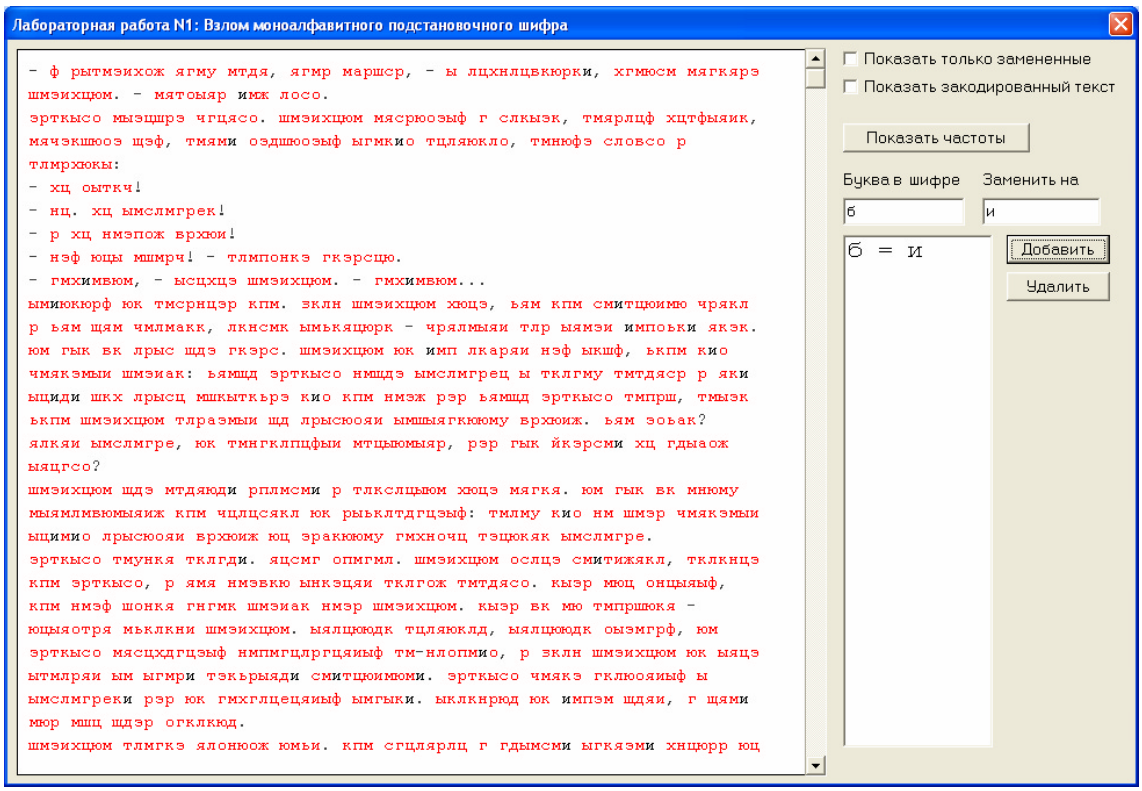

Рисунок 2. Изменения окна лабораторной работы после расшифровки одной буквы

Чтобы указать для какой-либо буквы шифра ее истинное (расшифрованное) значение, нужно в поле «Буква в шифре» указать значение буквы, например, “б”, а в поле «Заменить на» - ее истинное значение, например, “и”, а затем нажать кнопку “Добавить“. Результат такого действия приведен на рис. 2.

На рис. 3. Приведено окно выполнения лабораторной работы после добавления расшифровок нескольких букв.

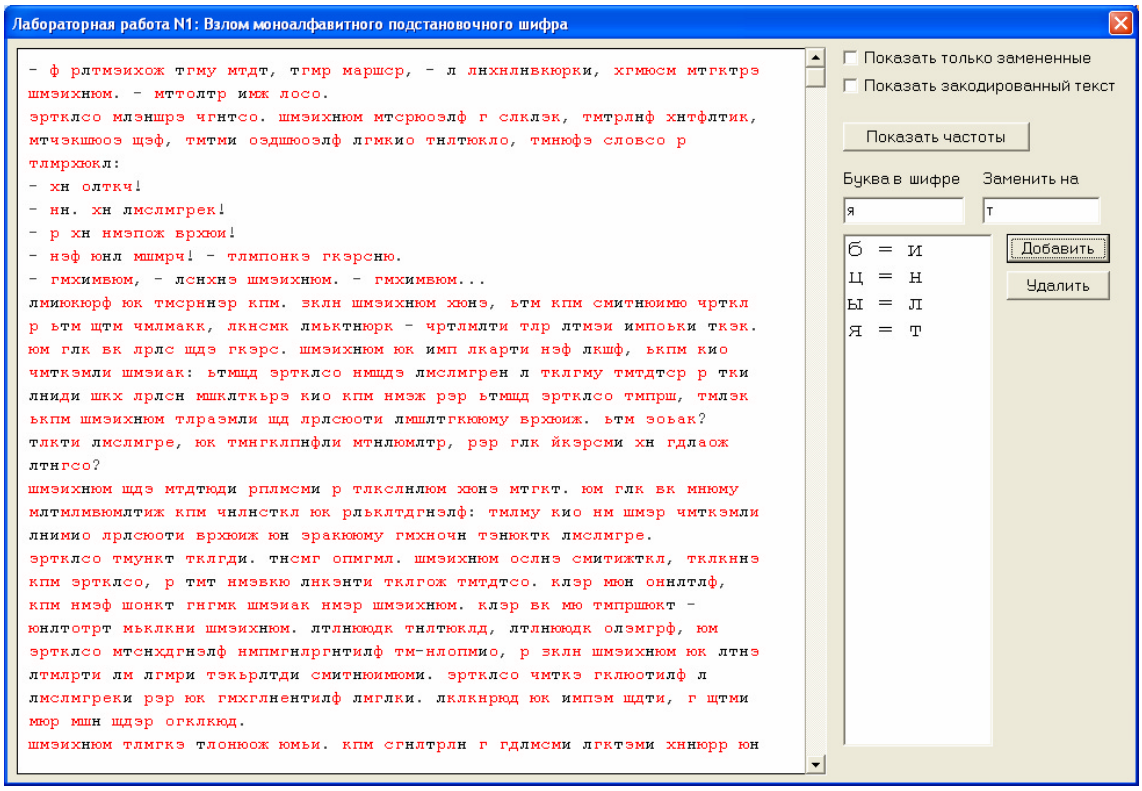

Рисунок 3. Окно лабораторной работы после расшифровки нескольких букв

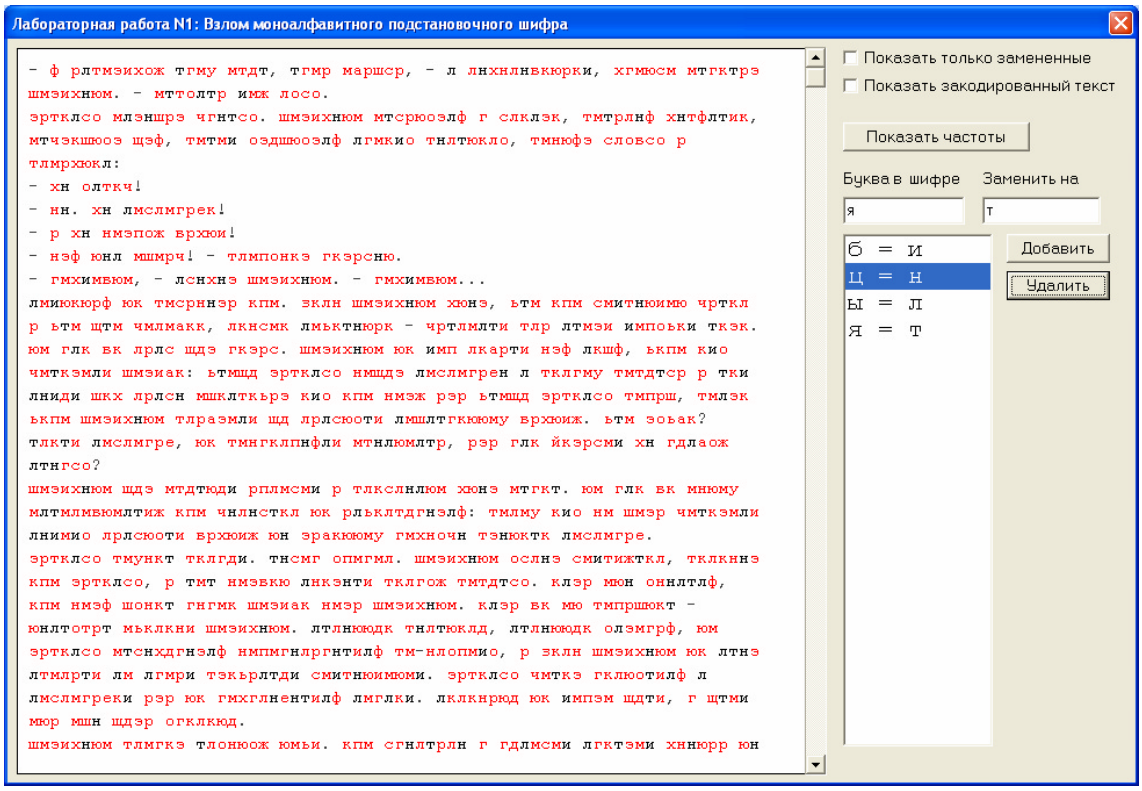

Рисунок 4. Процедура удаления ошибочно указанных расшифровок

Чтобы отменить указанную расшифровку буквы, нужно в списке расшифровок мышкой указать соответствующую пару букв и нажать кнопку «Удалить» (рис. 4).

Полоса вертикального скроллинга служит для навигации по расшифровываемому тексту.

2. Начинается частотная атака с анализа частот встречаемости букв в шифровке. Для этих целей в окне выполнения лабораторной работы предусмотрена кнопка «Показать частоты». При ее нажатии на экран выводится перечень десяти наиболее часто встречаемых букв в шифре, а также перечень букв, наиболее часто встречаемых в русском языке (рис. 5).

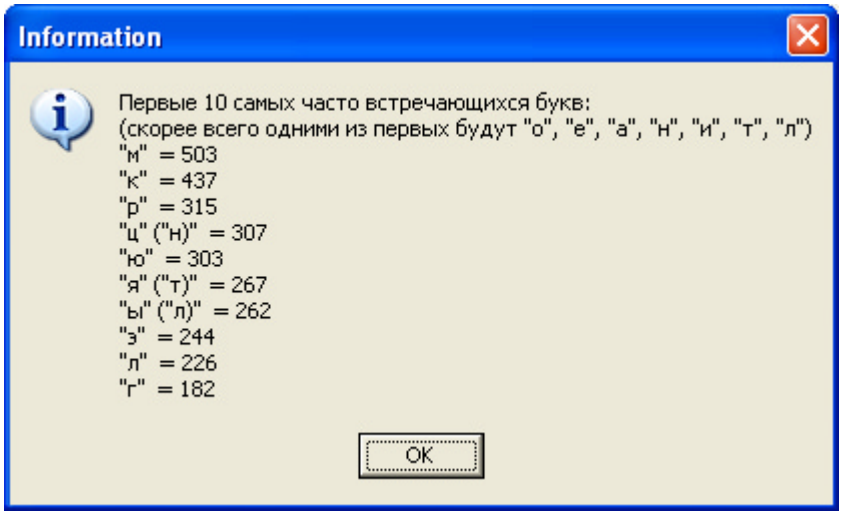

Рисунок 5. Информация о частотах встречаемости букв в шифре

Первым шагом в расшифровке текста может быть указание расшифровки для самой часто встречаемой буквы - буквы «о». Для случая, приведенного на рис. 5, указывается «о» как расшифровка буквы «м» шифра (см. рис. 6).

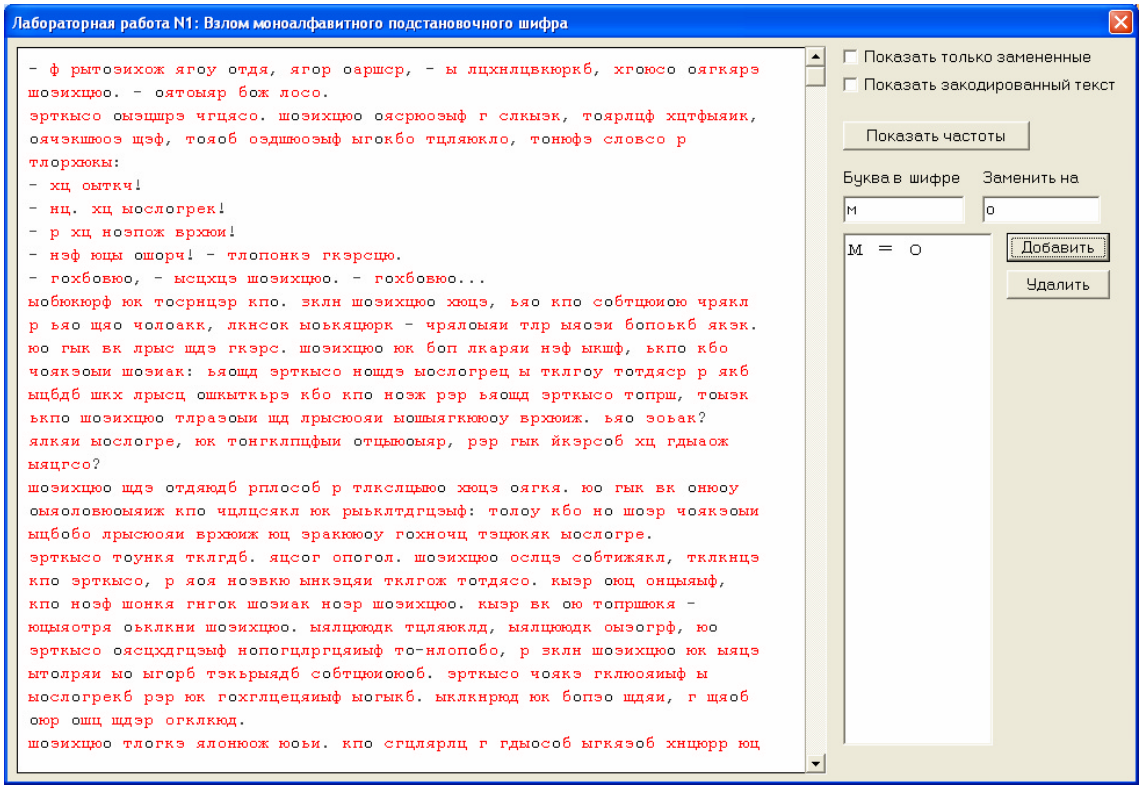

Рисунок 6. Первый шаг расшифровки - указание расшифровки буквы «о»

Следует помнить, что для конкретного текста частота встречаемости букв может быть несколько иной, чем в среднем для русского языка. Если в русском языке, например, буква

«т» встречается чаще, чем буква «л», то в каком-то конкретном тексте буква «л» вполне может встречаться чаще буквы «т». Поэтому слепо опираться на данные частотного анализа не следует.

3. В зашифрованном тексте осуществляется поиск коротких слов, зашифрованные буквы которых можно предсказать по уже расшифрованным буквам и частотной информации из рис. 5. На рис. 7. в верхней строчке есть фрагмент текста « ою », где «о» уже известно

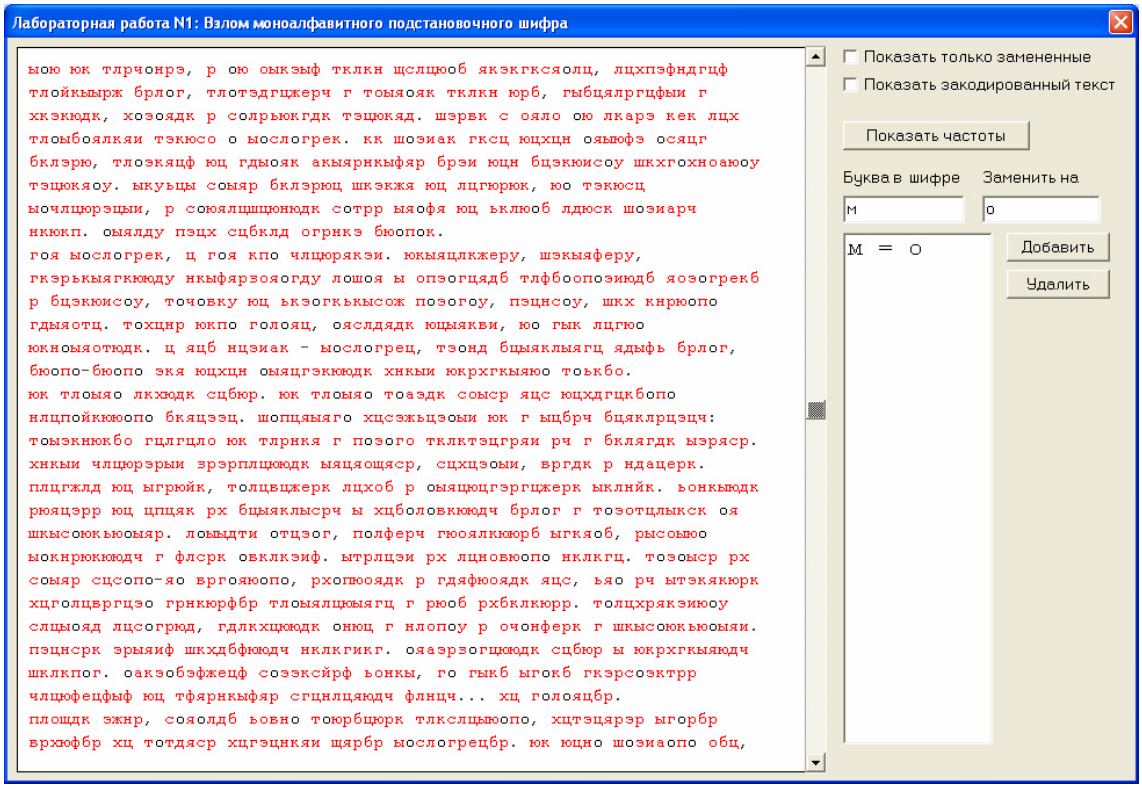

Рисунок 7. Поиск коротких слов

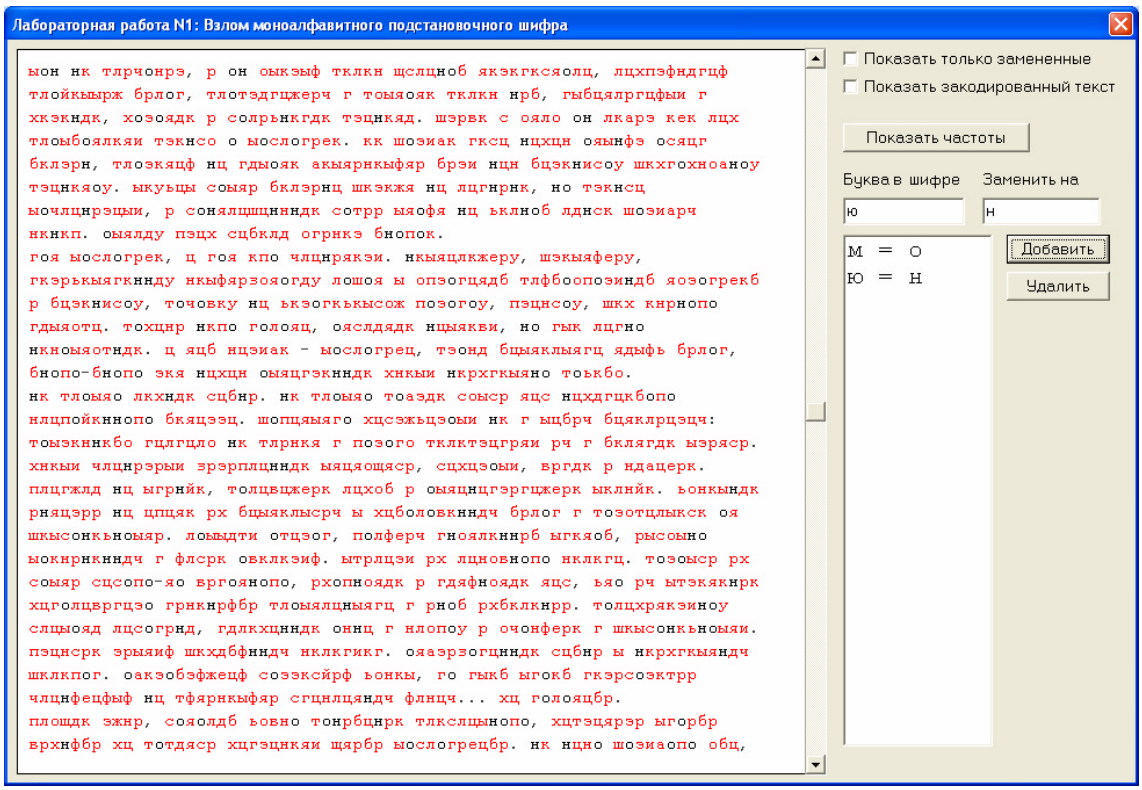

Рисунок 8. Результат расшифровки букв «о» и «н»

Этот фрагмент может быть скорее всего словом « он » В таблице частот (рис. 5) буква «ю» шифра стоит на 5-м месте, что примерно соответствует позиции буквы «н» русского языка (4-е место). Значит разумно попробовать поменять «ю» на «н». Результат приведен на рис. 8.

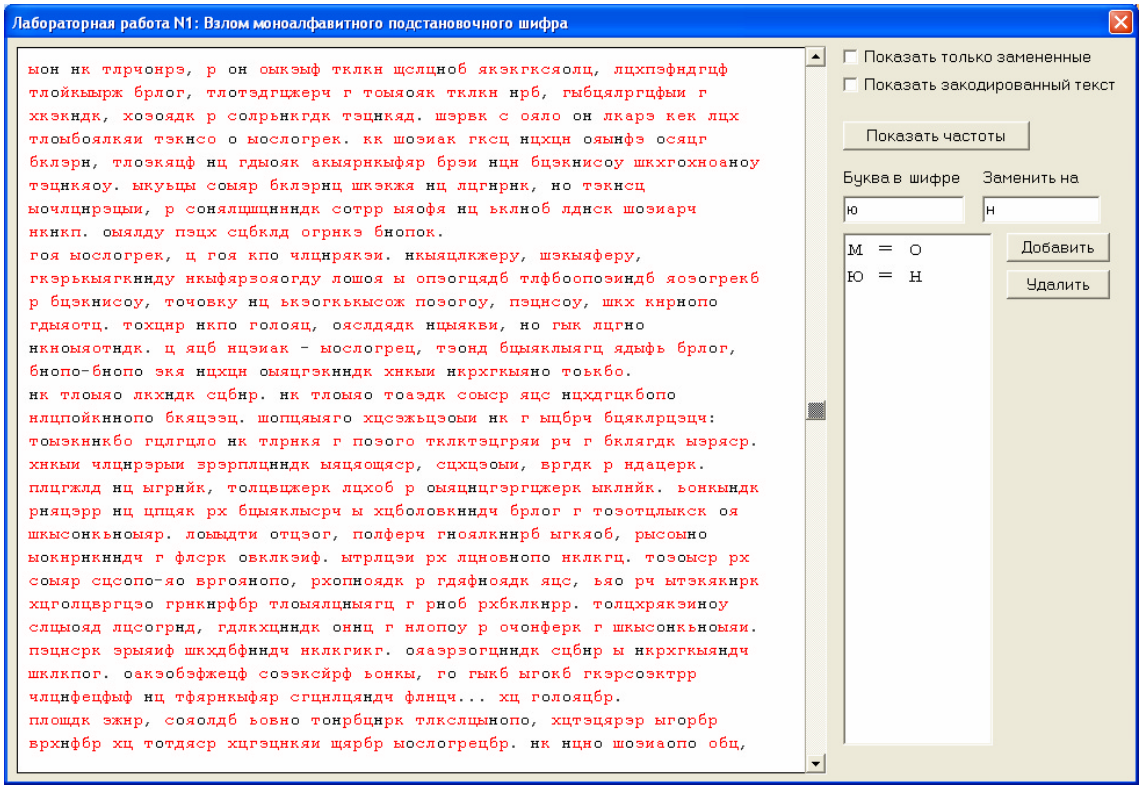

Рисунок 9. Продолжение поиска коротких понятных слов

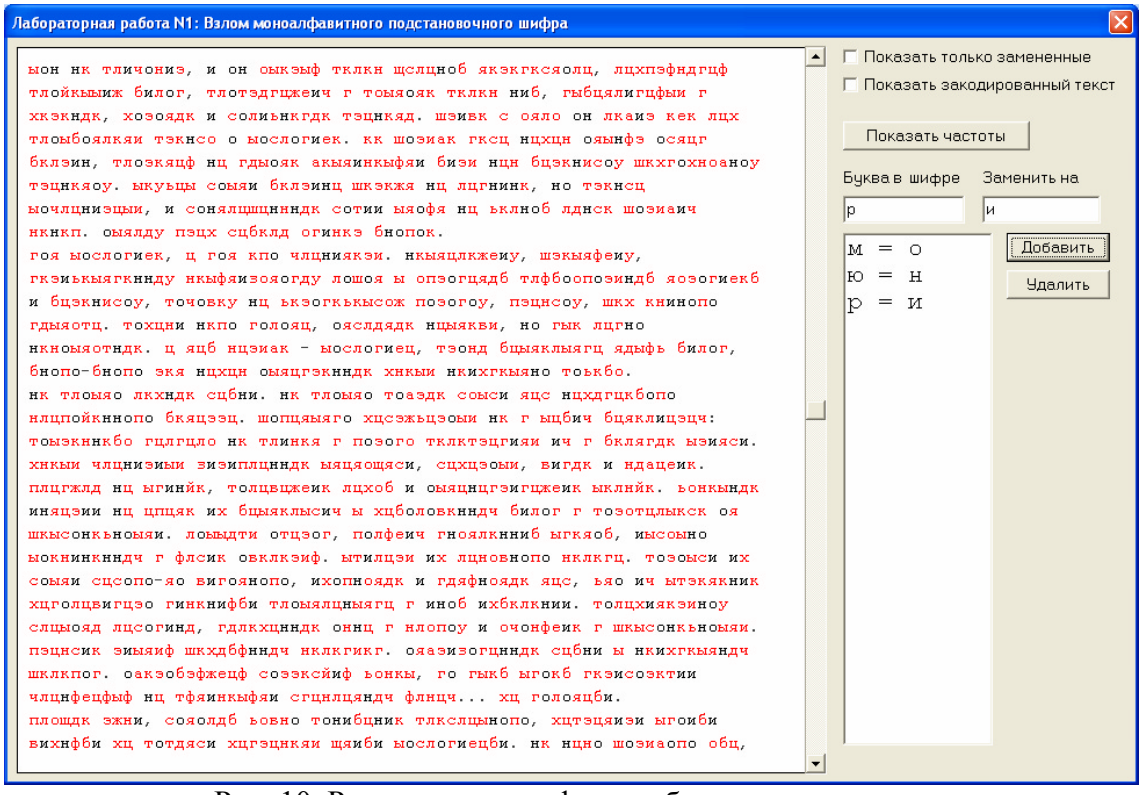

Рис. 10. Результат расшифровки букв «о», «н» и «и»

Далее повторяется поиск коротких слов, в которых можно догадаться о значении зашифрованных букв. На рис. 9 в первой и третей строках есть отдельно стоящее «р». Скорее всего это предлог «и», что согласуется и с информацией на рис. 5. Результат замены приведен на рис. 10.

На рис. 11 в первой строке обнаруживается слово из двух известных «и» и шифрованной буквы «э» между ними. Скорее всего это буква «л», образующая слово «или» (рис. 12).

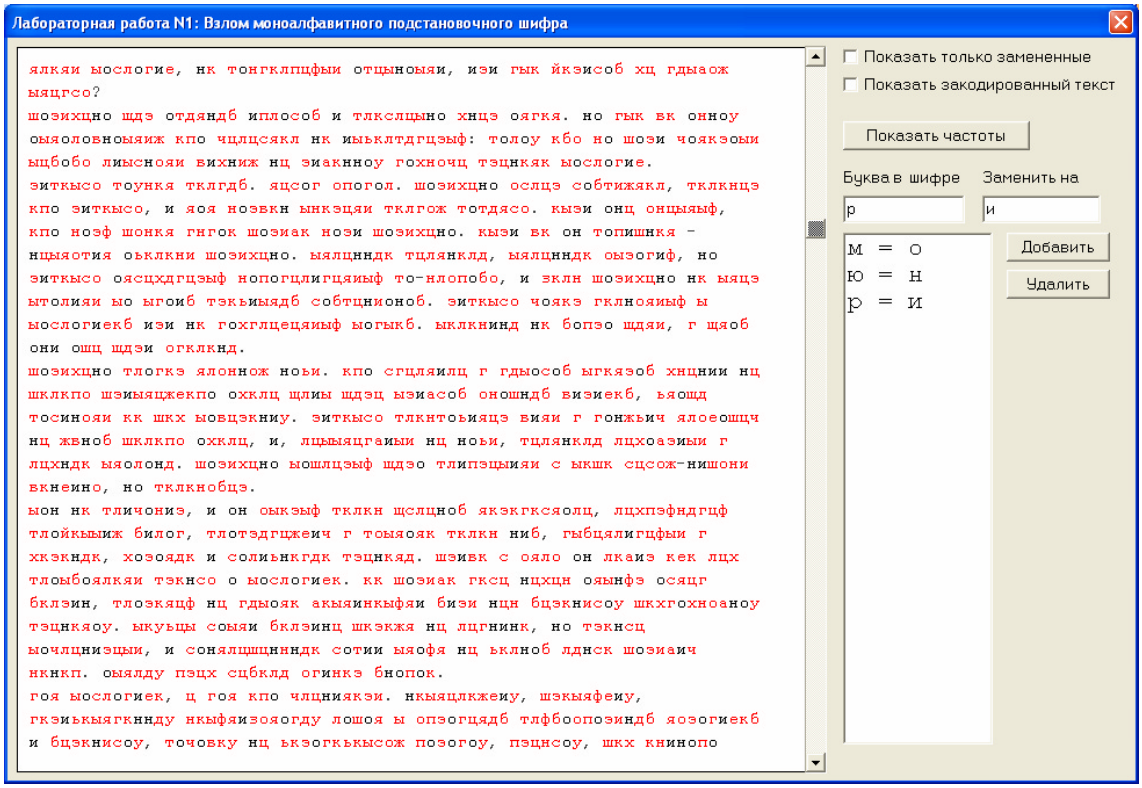

Рисунок 11. Продолжение поиска коротких понятных слов

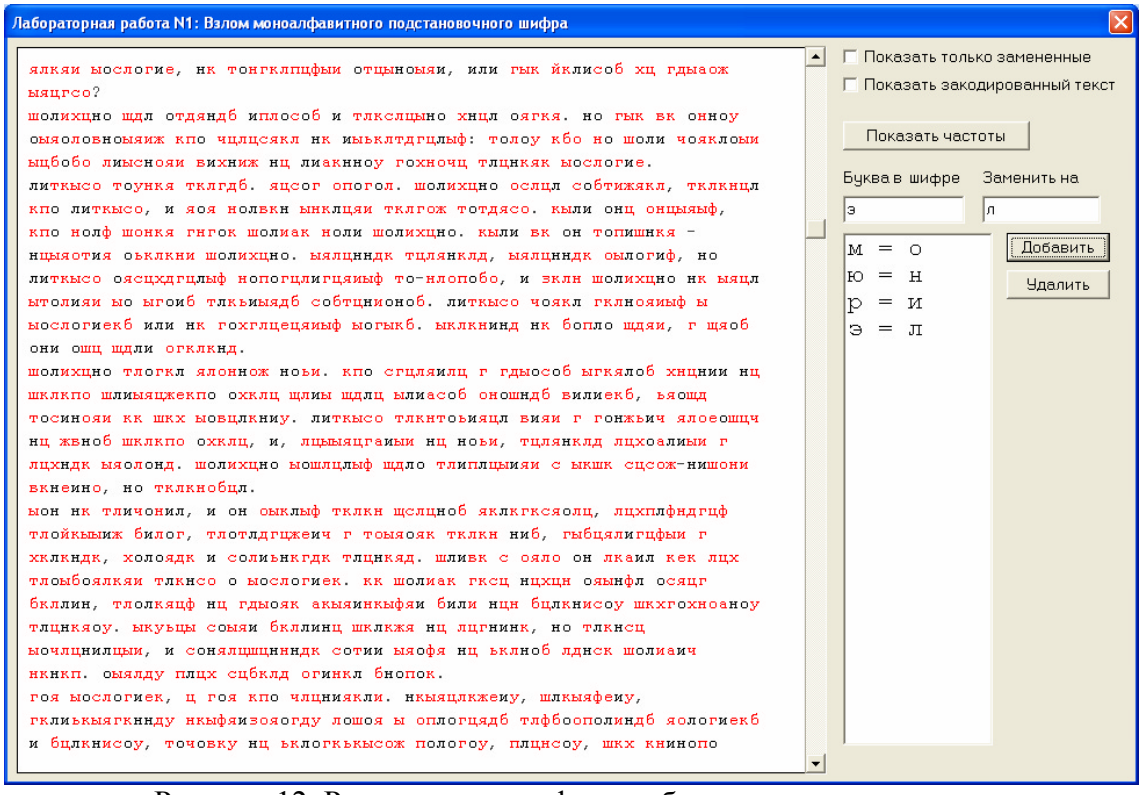

Рисунок 12. Результат расшифровки букв «о», «н», «и» и «л»

После расшифровки аналогичным образом букв «к» на «е», «ц» на «а» и «я» на «т» окно выполнения лабораторной работы приобретает следующий вид (рис. 13):

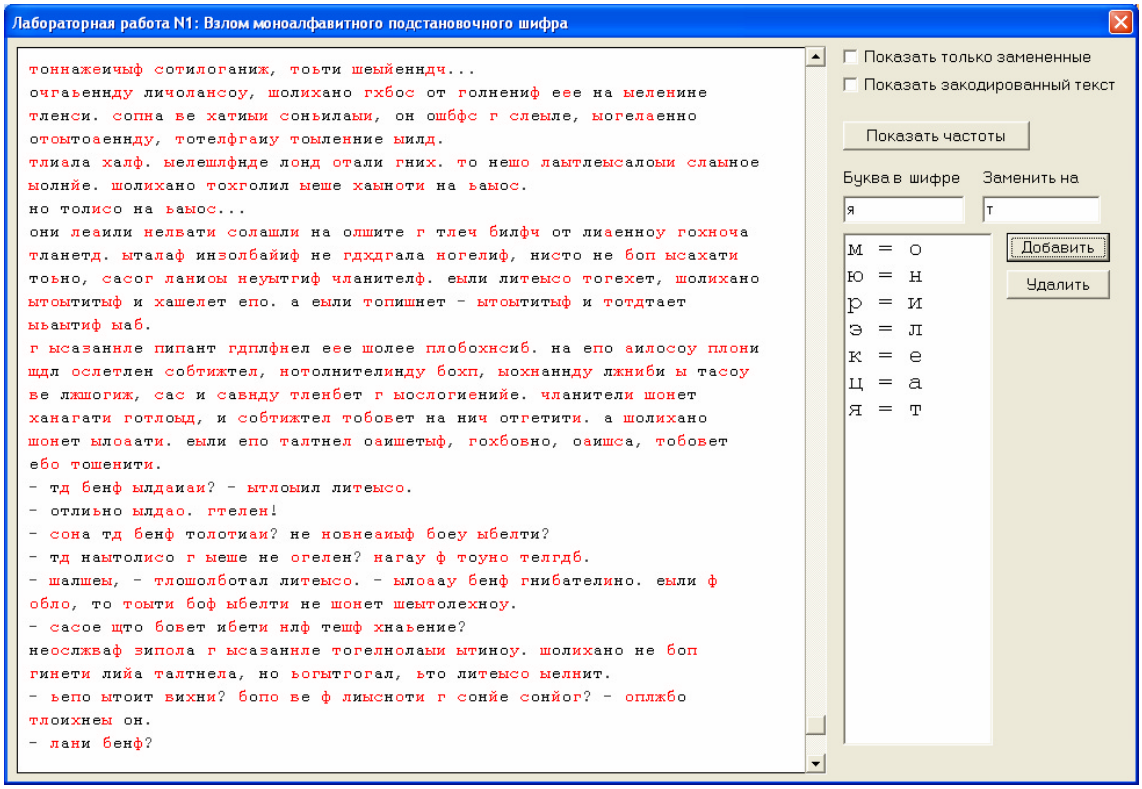

Рисунок 13. Окно выполнения лабораторной работы после расшифровки семи букв

Когда так много букв уже известно, зашифрованные буквы могут мешать для понимания слов. Для облегчения дальнейшего анализа в программе предусмотрена возможность выставления флага «Показать только замененные», при выставлении которого все зашифрованные буквы выводятся на экран в виде символов решетки (рис. 14).

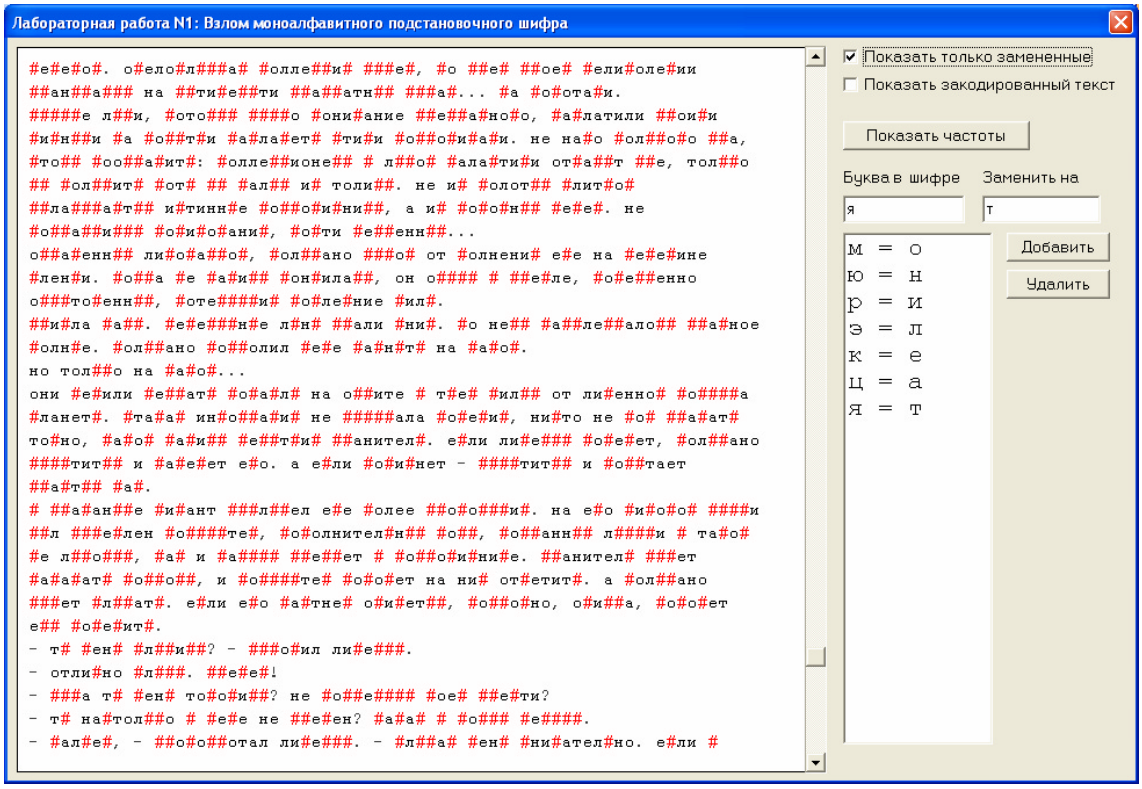

Рис. 14. Использование флага «Показать только замененные»

Теперь видно, что слово «##о#о##отал» в нижней строке вполне может быть словом «пробормотал». Если теперь выключить флаг, то можно получить косвенное подтверждение

этого - на позициях двух букв «р» в этом слове в шифре также находится одинаковая буква «л» (рис. 15).

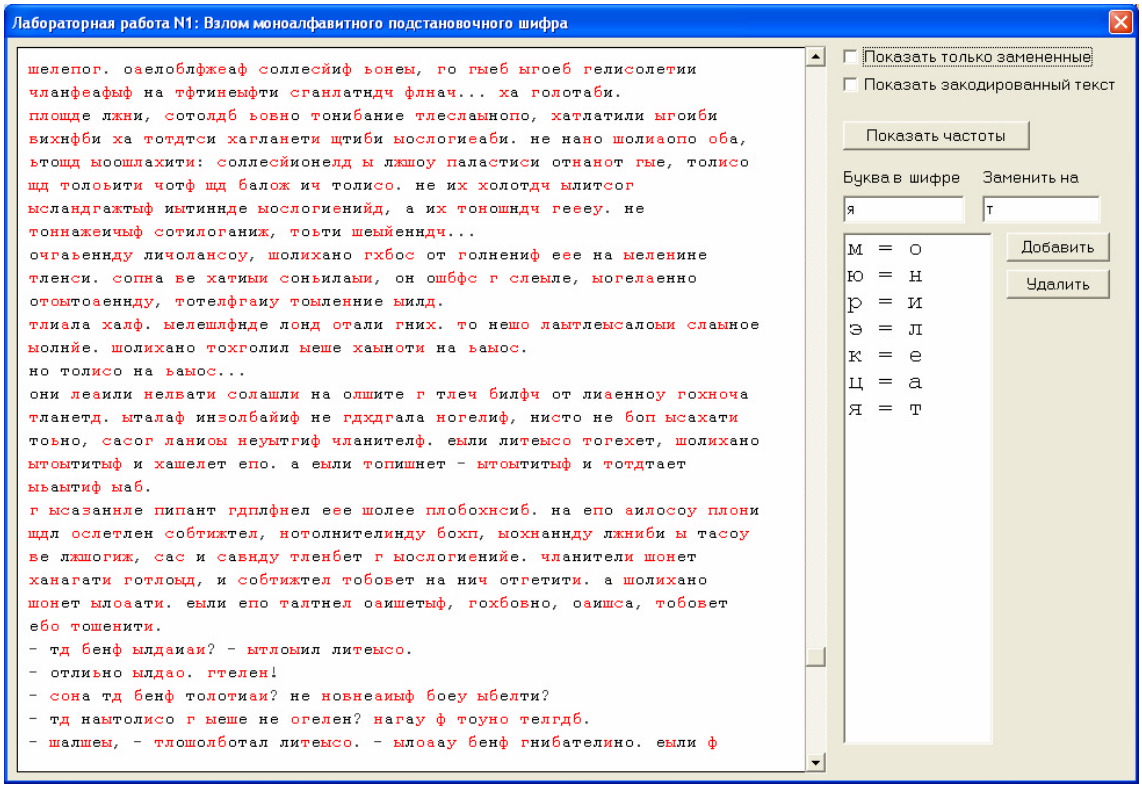

Рисунок 15. Проверка гипотезы отключением флага

Если заменить теперь букву «т» на «п», «л» на «р», «ш» на «б» и «б» на «м», то окно выполнения лабораторной работы станет выглядеть так(рис. 16):

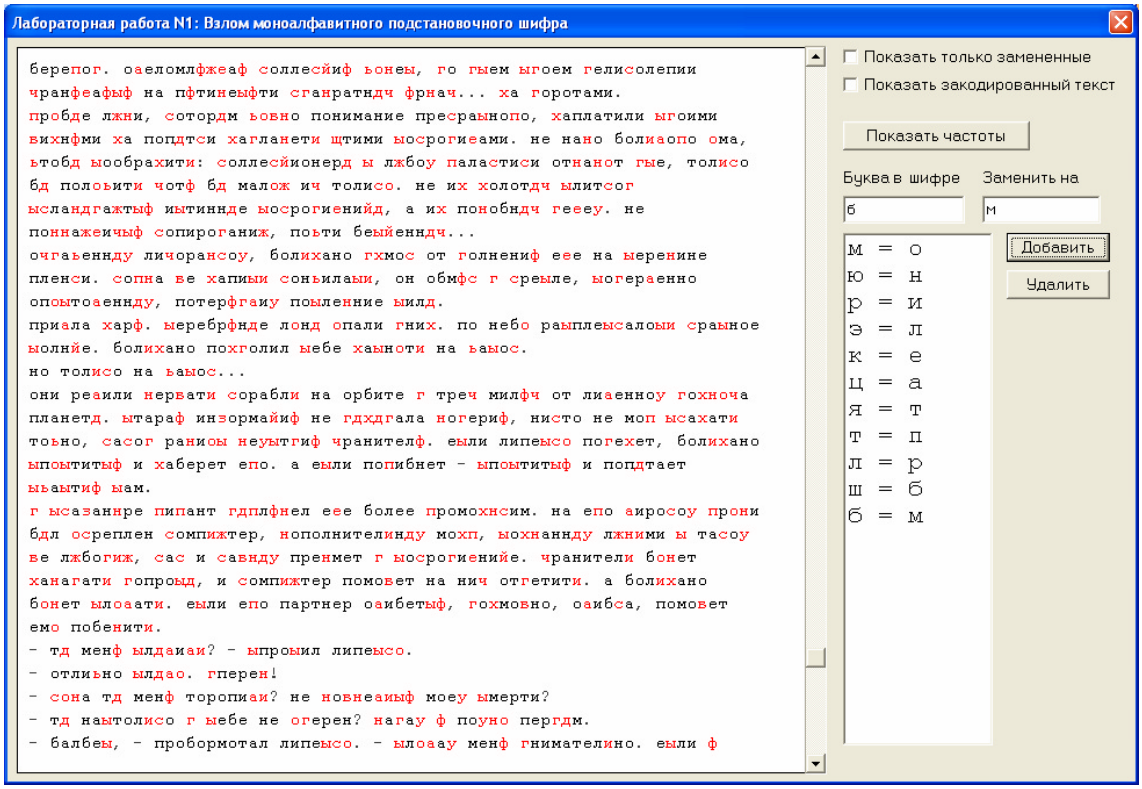

Рисунок 16. Окно лабораторной работы после расшифровки букв «п», «р», «б» и «м».

Хорошо видно, что дальнейший анализ значительно упрощается. Например, очевидно по слову «хаплатили», что буква «х» шифра соответствует букве «з» исходного текста. На рис. 17 приведено окно программы, когда анализ уже близок к завершению (осталось совсем немного нерасшифрованных букв).

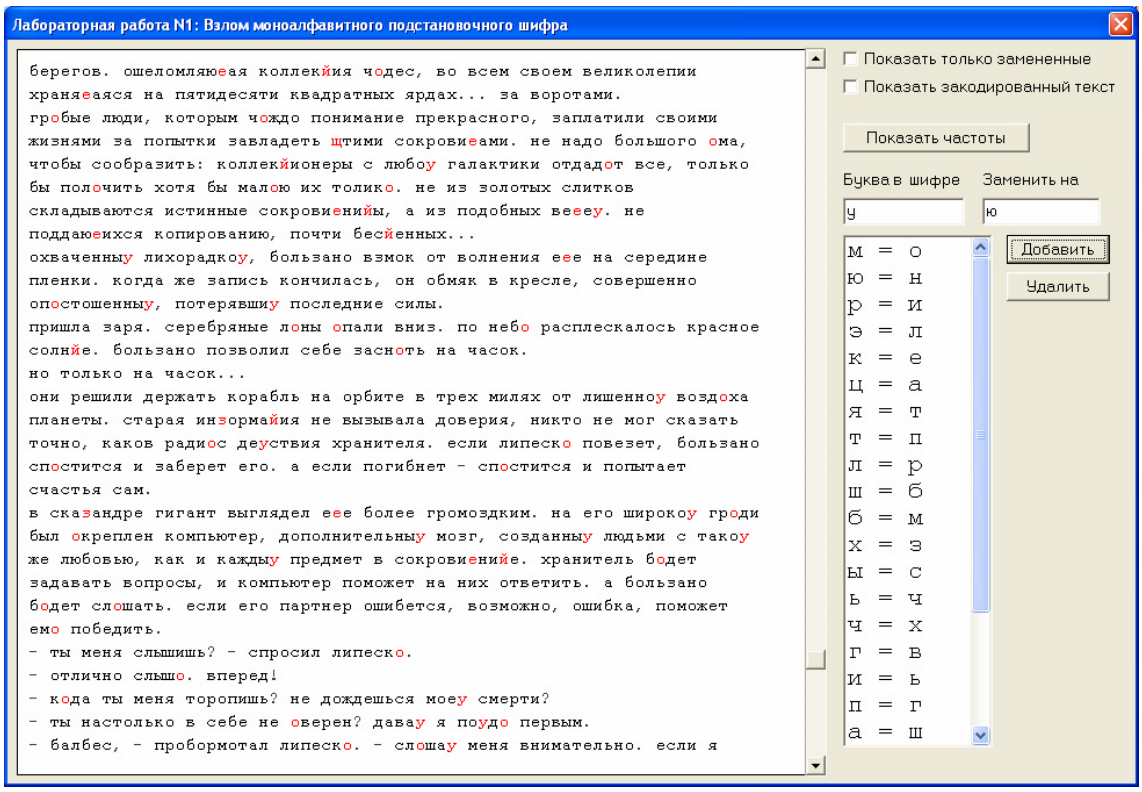

Рисунок 17. Расшифрованы почти все буквы текста

Когда же все буквы текста расшифрованы, на экран выводится информационное окно (рис. 18):

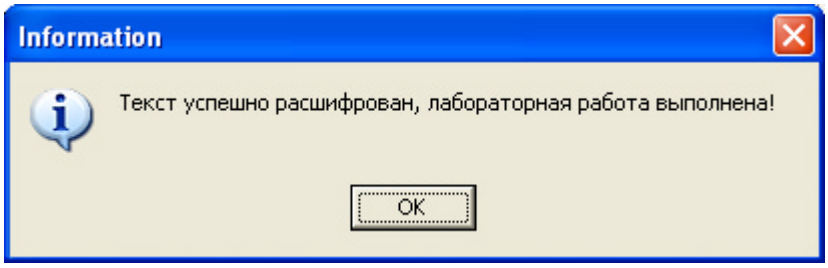

Рис. 18. Информационное окно, свидетельствующее о успешной расшифровке текста

Появление этого окна на экране свидетельствует об успешном выполнении лабораторной работы.