

Лабораторная работа

Частотные словари средствами MS Word и MS Excel

Задание 1: С помощью интернет-технологий найдите и сохраните в MS Word лирическую миниатюру Льва Ивановича Ошанина «Я номер набрал» (если задание выполняется не на уроке, возьмите **любое другое** стихотворение).

Задание 2: Преобразуйте текст стихотворения таким образом, чтобы каждое слово произведения начиналось в новом абзаце, т.е. его структура приобрела вид списка из слов. Для этого необходимо:

1. выделить текст стихотворения;
2. преобразовать все буквы к строчному виду (кнопка **Aa** или **Shift + F3**);
3. найти в тексте все **знаки препинания** и заменить их на **знак Абзаца**:
 - a. на вкладке **Главная**, выберите функцию **Найти** (или **Ctrl+N**);
 - b. в диалоговом окне **Найти и заменить** в поле **Параметры поиска** установите галочку **Подстановочные знаки**;
 - c. в строке **Найти** в квадратных скобках перечислите возможные знаки препинания в любом порядке **[.,,,:;'"()\!/? \]¹**;
 - d. в строке **Заменить на...** из списка **Специальный** выберите **знак Абзаца** (или **^p** впишите сами);
 - e. выберите функцию **Заменить все** (см. рисунок 1).
4. полученные в результате пустые абзацы можно убрать с помощью замены **каждых двух абзацев на один (^p^p на ^p)** – опция **Подстановочные знаки** при этом должна быть отключена, добейтесь полного отсутствия пустых строк.

Вопрос: Оцените объем стихотворения: сколько словоупотреблений встретилось в этом тексте?

Н.В.: выполняя шаг **с)**. 2-го пункта, обратите внимание на такой знак препинания, как «дефис». При его замене может произойти деление слова на две части, не являющимися самостоятельными словами. Например, слово «когда-нибудь».

¹ Не все знаки пунктуации можно перечислять в том виде, как они используются в предложении: некоторые из них сами являются командой для функции подстановочных знаков, поэтому их нужно перечислять через обратный слеш **\!** или **\?**

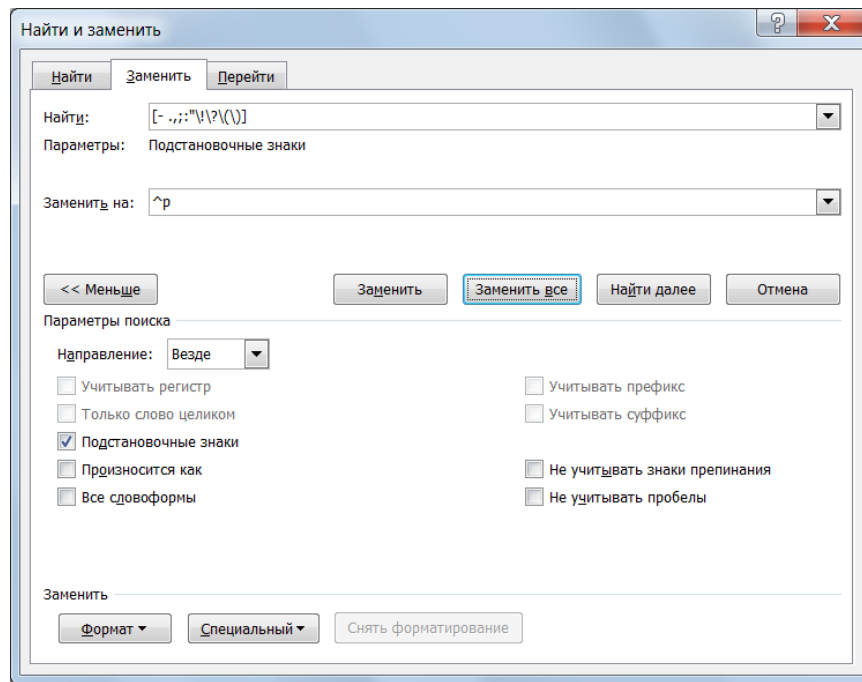


Рисунок 1

Вопрос: Что называют частотным словарем?

Задание 3: Постройте частотный словарь **грамматических категорий** выбранного стихотворения, оформив его в приложении MS Word. Для этого выполните следующие действия:

Предварительная работа:

1. выделите список слов стихотворения и перенесите в приложение MS Excel, в ячейку **A2** и затем продублируйте в ячейку **C2**;
2. сохраните документ, присвоив ему имя: «*Частотный словарь*»;
3. в ячейках **A1:C1** создайте шапку для будущей таблицы соответственно: «*Словоупотребления*», «*Грамматические категории*», «*Леммы*»;
4. в столбце **B** для каждого словоупотребления определите его грамматический признак (пользуйтесь автозаполнением, копированием, в спорных ситуациях проверяйте слово на gramota.ru);
5. в столбце **C** для каждого словоупотребления определите **лемму** (приведите каждое слово к его начальной форме), либо выполните это позже – в задании 4;
6. установите внутренние и внешние границы получившейся таблицы;
7. настройте функцию **Фильтр** для данной таблицы, позволяющий фильтровать список по категориям:
 - а. выделить ячейки **A1, B1** и **C1**;
 - б. на вкладке **Главная** в меню **Сортировка и фильтр** выбрать функцию **Фильтр**.

Создание словаря:

1. создайте новый документ в MS Word «Частотный словарь грамматических категорий»;
2. в документе «Частотный словарь грамматических категорий» создайте таблицу: 3 столбца на 3 строки;
3. оформите шапку таблицы, присвоив столбцам следующие наименования соответственно: «*Грамматические категории*», «*Словоупотребления*», «*Частота употребления слов*»;
4. в столбце «*Грамматические категории*» перечислите грамматические категории: *существительное, прилагательное* и т.д. (если необходимо, добавляйте строки);
5. в столбце «*Словоупотребления*» напротив каждой категории перечислите все словоупотребления с данным признаком:
 - a. перейдите в документ **Частотный словарь.xlsx**;
 - b. используя функцию **Фильтр** в столбце **В**, отберите только строки с грамматической категорией **Глагол**;
 - c. выделите и скопируйте полученный список словоупотреблений;
 - d. вставьте скопированный список в таблицу MS Word напротив грамматической категории **Глагол**: для этого используйте в меню **Вставить** вкладки **Главная**, функцию **Специальная вставка... – Неформатированный текст**;
 - e. используя функцию **Найти и заменить** (Ctrl+N), преобразуйте список слов в словарный ряд, заменив знак **Абзаца** на такую комбинацию знаков, как «**запятая-пробел**» (см. рисунок 2);
 - f. проделайте те же самые действия и для оставшихся категорий.
6. в столбце «*Частота употребления слов*» напротив грамматической категории **Глагол** укажите количество словоупотреблений с данным признаком;
7. проделайте те же самые действия и для остальных категорий;
8. частотный словарь грамматических категорий готов.

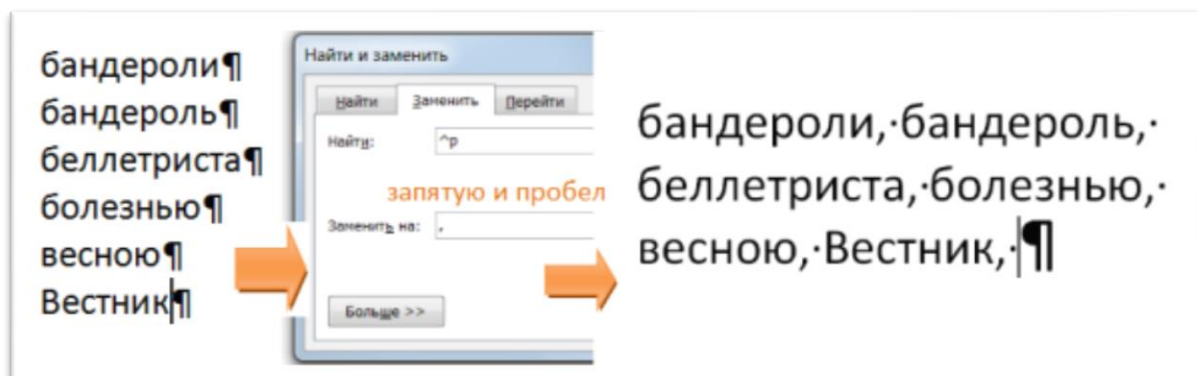


Рисунок 2

Вопрос: Что называют леммой?

Задание 4: Постройте лемматизированный частотный словарь выбранного стихотворения, оформив его в приложении MS Excel. Для этого в документе **Частотный словарь.xlsx** выполните следующее:

1. в столбце **С** для каждого словоупотребления определите его лемму (если не выполнили этого ранее, в задании 3);
2. выделите столбцы **В** и **С**;
3. на вкладке **Вставка** в меню **Сводная таблица** выбираем функцию **Сводная таблица**;
4. на новом листе документа справа от рабочего поля документа активизируется дополнительное поле: **Список полей сводной таблицы**;
5. в области **Выберите поля для добавления в отчет** установите галочку **Леммы**;
6. вы получаете список уникальных слов стихотворения;
7. отобразите частоту употребления для каждой леммы в отдельности:
 - а. наведите курсор мыши на функцию **Леммы** в области **Выберите поля для добавления в отчет**;
 - б. удерживая левую кнопку мыши, перетащите выделенную функцию в область **Σ Значение** (см. рисунок 3, шаг 1);
8. отсортируйте лемматизированный частотный словарь по убыванию: стоя на ячейке **В4** (первая ячейка с числом), на вкладке **Данные** выберите функцию **Сортировка** по убыванию (от Я до А);
9. установите фильтр отчета по грамматическим категориям:
 - а. наведите курсор мыши на функцию **Грам. категории** в области **Выберите поля для добавления в отчет**;
 - б. удерживая левую кнопку мыши, перетащите выделенную функцию в область **Фильтр отчета** (см. рисунок 3, шаг 2);
10. теперь, если требуется, в ячейке **В1** с помощью фильтра можно отбирать леммы по грамматическим категориям.

Вопросы:

1. Сколько всего словоупотреблений в стихотворении?
2. Сколько уникальных слов встречается в стихотворении?
3. Каков процент употребления уникальных слов в стихотворении по отношению ко всем словам?

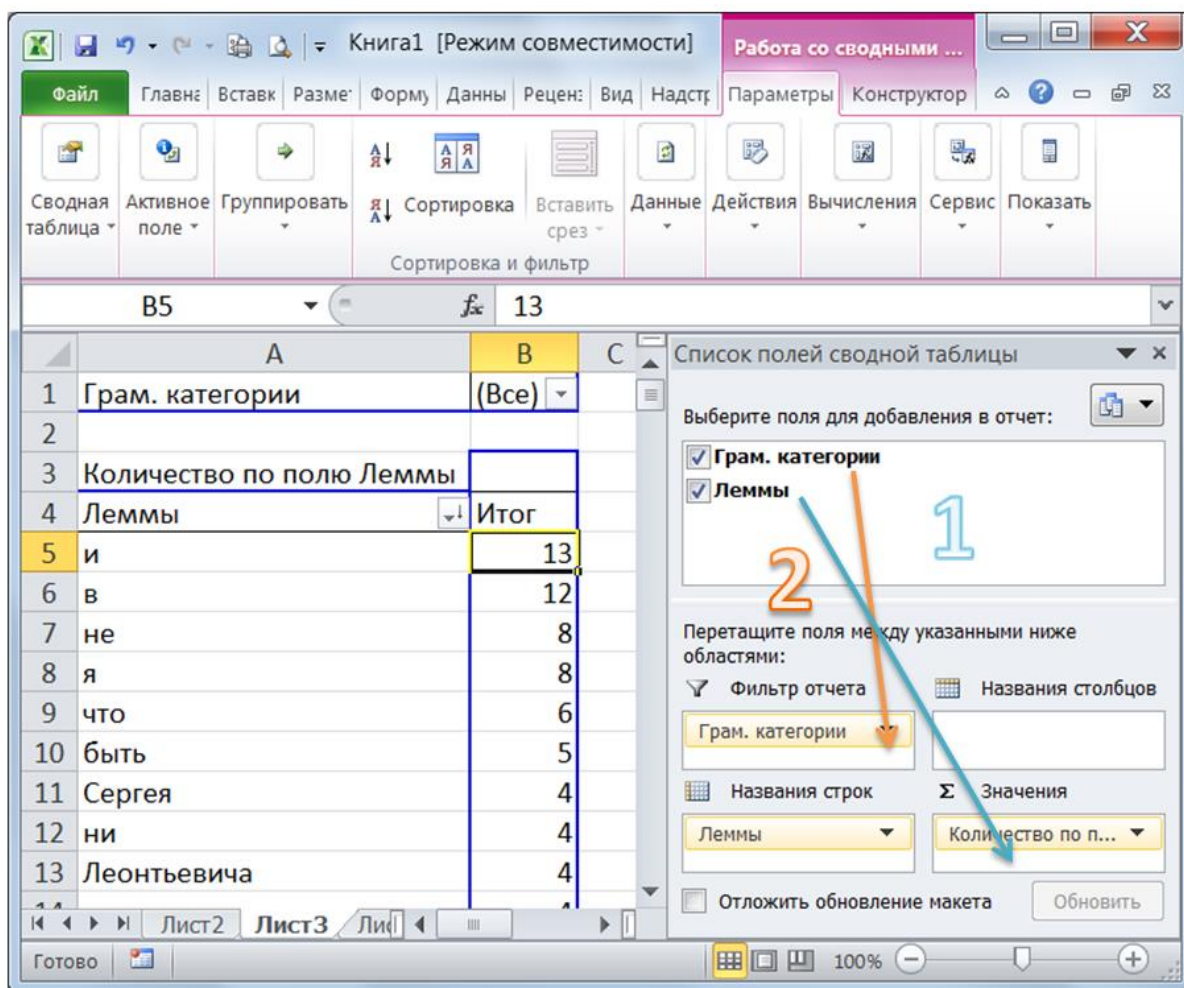


Рисунок 3

Дополнение таблицы в MS Word:

1. откройте документ Частотный словарь грамматических категорий.docx;
2. смените ориентацию страницы на Альбомная;
3. в таблице после столбца «Словоупотребления» добавьте столбец «Леммы»;
4. в столбце «Леммы» напротив каждой категории перечислите все леммы с данным признаком:
 - а. перейдите в документ **Частотный словарь.xlsx** на вкладку со сводной таблицей;
 - б. используя функцию **Фильтр отчета** в столбце **В1**, отберите только строки с грамматической категорией **Глагол**;
 - в. выделите и скопируйте полученный список лемм вместе с количеством их употреблений;
 - г. вставьте скопированный список в таблицу MS Word напротив грамматической категории **Глагол**: для этого используйте в меню **Вставить** вкладки **Главная**, функцию **Специальная вставка... – Неформатированный текст**;

- e. используя функцию **Найти и заменить** (Ctrl+N), преобразуйте список слов в словарный ряд, заменив знак **Абзаца** (^p) на такую комбинацию знаков, как «запятая-пробел» (см. рисунок 2);
 - f. затем, используя функцию **Найти и заменить** (Ctrl+N), замените знак **Табуляции** (^t) на комбинацию знаков «неразрывный пробел», «неразрывный дефис» и снова «неразрывный пробел» (^s^~^s), чтобы указанное для каждой леммы число всегда было рядом с леммой, к которой оно относится (см. рисунок 4);
 - g. сделайте те же действия и для оставшихся категорий.
5. придайте таблице завершённый вид.

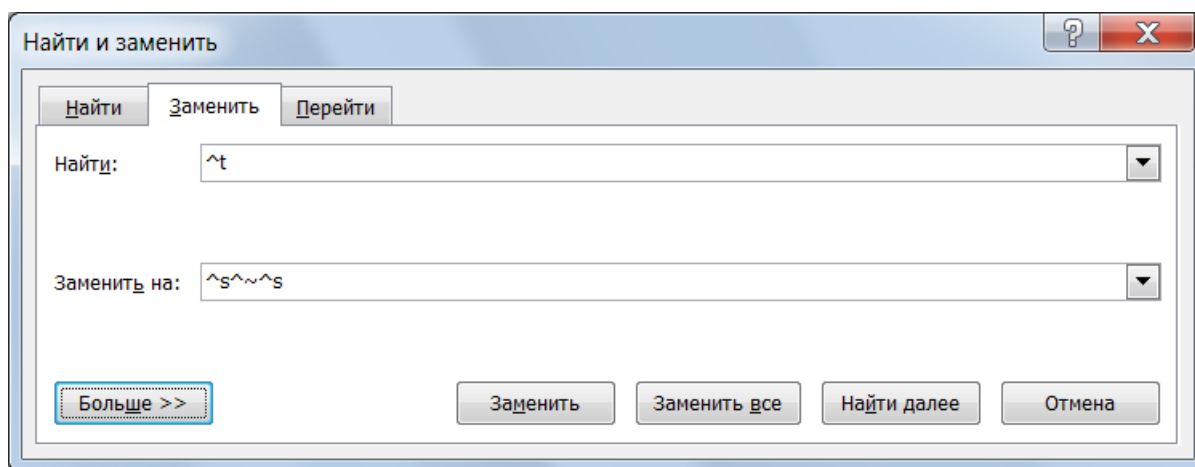


Рисунок 4

Факультативно:

Аналогичные операции можно проводить и **для построения частот букв в тексте**. Для этого в задании 2 вместо диапазона знаков пунктуации в поле «Найти» следует указать в специальных командах «**Любой знак**» (или ^?), а в поле «Заменить на» выбрать «**Искомый текст**» (или ^&) и «**Знак Абзаца**» (или ^p). Опция «Подстановочные знаки» при этом должна быть **отключена** (см. рис. 5). Все знаки распределятся на отдельных строках, при сортировке по возрастанию верхние строки (до букв) можно будет легко удалить.

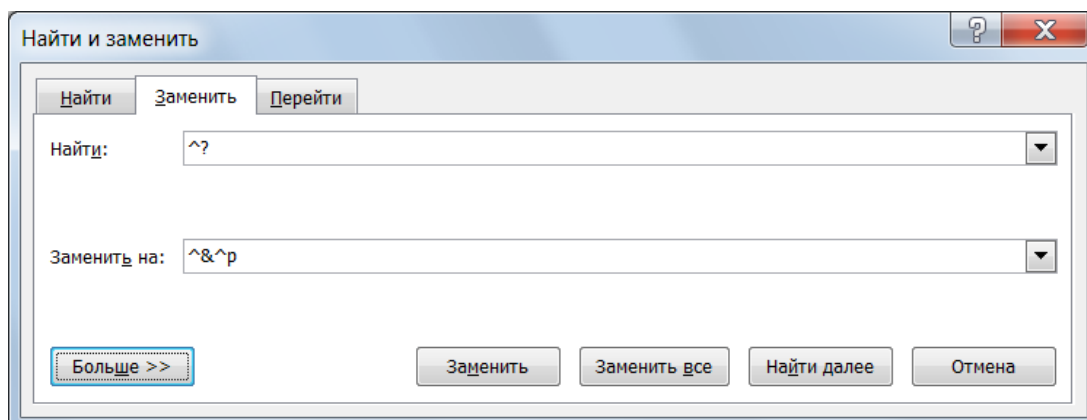


Рисунок 5