## Лабораторная работа 2.1.

Исследовать текст

«Ромео и Джульетта» В. Шекспира

в переводе Б. Пастернака и Т. Щепкиной-Куперник

(теоретико-множественный подход к моделированию

лингвистических процессов)

#### Подготовительная работа

- 1. Откройте перевод <u>Б. Пастернака</u> и скопируйте из Акта II сцену 2 (со слов «Сад Капулетти»)
- Создайте Текстовый документ в своей папке с именем Пастернак.txt и вставьте в него скопированный текст.
- Откройте перевод <u>Т. Щепкиной-Куперник</u> и скопируйте из Акта II сцену 2 (со слов «Сад Капулетти»)
- 4. Создайте **Текстовый документ** в своей папке с именем **Щепкина-Куперник.txt** и вставьте в него скопированный текст.
- 5. Запустите программу WordTabulator.

# Загрузка файлов в WordTabulator



WordTabulator v3.5 %temp%\wt\$proj.wt	-		$\times$
Исх.файлы Настройки Обработка Слов Обработка Тегов Выходной Индекс Помощь			
Фильтр для файлов: *.txt П П П С С С С С С С С С С С С С С С С			
— Множество Входных Текстов			_
<ul> <li>D:\YandexDisk\Преподавание\ОМОИ\Пастернак.txt</li> </ul>			
<ul> <li>D:\YandexDisk\Преподавание\ОМОИ\Щепкина-Куперник.txt</li> </ul>			
<ul> <li>Множество Исключений</li> <li>1. Чтобы видеть txt файлы, смените ф</li> <li>2. Станьте на "Множество входных текси и нажмите кнопку загрузки файлов.</li> <li>3. Убедитесь, что нужные файлы загрураскрыв (+) слева от папки.</li> <li>4. Нажмите кнопку с зеленым человечи раскрыв (+) слева от папки.</li> <li>4. Нажмите кнопку с зеленым человечи начать построение частотного словаря бело элементов: 729 (3508 байтов)</li> <li>5. Посмотрите на результат, нажав на '</li> </ul>	иль стов /жен (ом,	тр 3" ны, что и".	бы
Формирую выходной индекс Затрачено времени: 140 msec			~
Затрачено времени 0:00:00			

#### Настройка словаря – Выходной индекс

₩ WordTabulator v3.5 %temp%\wt\$proj.wt	<
Исх.файлы Настройки Обработка Слов Обработка Тегов Выходной Индекс Помощь	
Выходной индекс \Nutik\AppData\Local\Temp\wt\$index.html Формат индекса HTML • Вывести элементы все • иастоте • убыв. •	
Пост-Обработка индекса Править	

На вкладке "Выходной индекс" можно настроить вид частотного словаря:
1) сортировать по "частоте".
2) по убыванию (от большего к меньшему)

## Множество исключений

WordTabulator v3.5 %temp%\wt\$proj.wt	-		×
Исх.файлы   Настройки   Обработка Слов   Обработка Тегов   Выходной Индекс	Помоц	ць	
Фильтр для файлов: *.txt • • • • • • • • • • • • • • • • • •	5TOP	26	1
<ul> <li>Множество Входных Текстов</li> <li>D:\YandexDisk\Преподавание\ОМОИ\Пастернак.txt</li> <li>Множество Исключений</li> <li>D:\YandexDisk\Преподавание\ОМОИ\Щепкина-Куперник.txt</li> </ul>			

 Перетащите файл с текстом Щепкиной-Куперник из Множества Входных Текстов в папку Множество Исключений, в этом случае можно будет настроить взаимодействие между этими двумя папками

## Настройка – Обработка исключений





- 1. После указанных настроек нажмите на обработку (зеленый человечек) и на просмотр (очки).
- 2. Выделите весь текст командой **Ctrl+A** и скопируйте его (Ctrl+C).
- 3. Создайте в папке файл RTF с именем Лабораторная по Шекспиру.rtf и вставьте скопированный текст на чистый лист (Ctrl+V).
- 4. Переименуйте строку «Тип элементов: словоформа»
   в «Общая лексика (А /\ В)»



#### Настройка исключений – Вычитание

WordTabulator v3.5 %temp%\wt\$proj.wt	- 🗆 X
Исх.файлы Настройки Обработка Слов Обрабо	тка Тегов   Выходной Индекс   Помощь
Поисковые шаблоны	Кодировка текстов Windows-1251 Тип элементов
↓	словоформа. 💌
🗁 регулярные выражения	1 💌 минимальная длина слова
🔲 учитывать регистр букв	
🗌 использовать морфологический модуль	→ минимальная частота
Символы разделители .,!?:+=V"(){}[]	Обработка исключений вычитание
Дополнительные буквы	
Дополнительные символы	

• Смените обработку исключений на «Вычитание»

# Вычитание (уникальная лексика)

- 1. После указанных настроек нажмите на обработку (зеленый человечек) и на просмотр (очки).
- 2. Выделите весь текст командой **Ctrl+A** и скопируйте его (Ctrl+C).
- 3. Перейдите в конец лабораторной командой **Ctrl+End** и нажмите **Ctrl+Enter**, чтобы начать новую страницу, затем вставьте скопированный текст (Ctrl+V).
- 4. Переименуйте строку «Тип элементов: *словоформа»* в «**Уникальная лексика Б. Пастернака (А \ В)**»





🗹 WordTabulator v3.5 %temp%\wt\$proj.wt	_		×
Исх.файлы   Настройки   Обработка Слов   Обработка Тегов   Выходной Индекс	Помоц	ць	
Фильтр для файлов: *.txt <ul> <li>Фильтр для файлов:</li> <li>Ф. Схать файлы в поддиректориях</li> </ul>	STOP	2	3
<ul> <li>Множество Входных Текстов</li> <li>D:\YandexDisk\Преподавание\ОМОИ\Щепкина-Куперник.txt</li> <li>Множество Исключений</li> <li>D:\YandexDisk\Преподавание\ОМОИ\Пастернак.txt</li> </ul>			

 Перетащите файл с текстом Щепкиной-Куперник назад в папку Множество Входных Текстов, а файл с текстом Пастернака переместите в папку Множество Исключений, обработку слов оставьте той же (Вычитание).

# Вычитание (уникальная лексика)

- 1. После указанных настроек нажмите на обработку (зеленый человечек) и на просмотр (очки).
- 2. Выделите весь текст командой **Ctrl+A** и скопируйте его (Ctrl+C).
- 3. Перейдите в конец лабораторной командой **Ctrl+End** и нажмите **Ctrl+Enter**, чтобы начать новую страницу, затем вставьте скопированный текст (Ctrl+V).
- 4. Переименуйте строку «Тип элементов: *словоформа»* в «Уникальная лексика Т. Щепкиной-Куперник (В \ А)»



# Обработка таблиц

- 1. Программу WordTabulator можно закрыть.
- 2. Отредактируйте все три полученные таблицы по следующим шагам:
  - Удалите столбцы # и «Ссылки», которые дублируют информацию частотных столбцов и «раздувают» размер файла своими гиперссылками.

т	Co	ылки	Рассылки	Рецензир	ование Вид	ABBYY FineRead	er 12 Конструктор	р Маке	т 🗘 Помоц	цн Я
ц	L 1	Удалит Удалит Т У У У У У У У У У У У У У	сверху /далить <u>я</u> чей /далить стол /далить стро	Встан Встан Ки I бцыо ки ицу	зить снизу вить справа гъ 1 + 2 + 1 + 3 +	<ul> <li>Объединить и</li> <li>Разделить ячи</li> <li>Разделить таб</li> <li>Объединени</li> <li>4 · 1 · 5 · 1</li> </ul>	ейки с. 0,48 ст. 3,54 ст. 5лицу ∺ Автопо 9 Размер я 6 · 1 · 7 · 1 · 8 · 1 ·	м ‡ ⊞ м ‡ ⊞ одборт чейки Га 9 <u>∄</u> т • 10 •	Выравнивание •	Даннь •
			+++			Оощая-л	ексика (АЛ.В)			-
	-			#α	Элемента		🤌 Ссылкиа	τ	Iастота¤	0
	2 .			1.¤	джульетта	α	<u>1</u> (10)¤	10¤		o
				2.¤	<b>и</b> Ω		<u>1</u> (10)¤	10¤		0
				3.¤	ромеоа		<u>1</u> (9)¤	<b>9</b> ¤		0
								-		i.



 – Добавьте справа столбец, в который позже будут вставлены слова, имеющие частоту в 1 употребление.

Ссылки	Рассылки	Рецензирование	Вид ABBYY Fir	neReader 12	Конструктор	Макет	🛛 Помощн	+ <i>P</i> _	두
Удалит •	вставит сверху Строки и	Вставить сни: Вставить слег Вставить спри вставить спри	зу 🔂 Объеди за Раздел ава Раздел объе	инить ячейки ить ячейки мв таблицу динение	<ul> <li>0,48 см</li> <li>2,79 см</li> <li>Автоподбо</li> <li>Размер ячейк</li> </ul>	⇒	равнивание /	Данные Т	
			Оби	цая лексик	a (A /\ B)	.11 . 1 . 12 .	1 . 13 . 1 . 14 .	. 1 . 12 . 1	10
4 •	Элеме	ент	Частота		Част	ота: 1			
	лжуль	etta 10	)						
t	и	10							
1	ромео	9							
÷	не	7							

- Очистите столбец «Частота» в тех ячейках, где указаны только единицы.
- Отсортируйте по алфавиту слова, употребленные по 1 разу (выделить ячейки с ними и нажать 🗍 )

# Обработка таблиц - 3

- Объедините ячейки со словами, которые употреблены единожды в одну большую ячейку.
- Объедините ячейки столбца «Частота: 1»
   тоже в одну большую ячейку (но до начала нижней большой ячейки).
- Перенесите
   содержимое нижней
   большой ячейки в
   верхнюю правую.





 – Удалите образовавшиеся пустые строки (только не удаляйте последнюю строку с общим количеством)



# Обработка таблиц - 5

Рег	цензирование	Вид	ABB	YY FineReader 12	2	Конструктор	Макет
низу лева права Гя	<ul> <li>Объединит</li> <li>Разделить я</li> <li>Разделить т</li> <li>Объедине</li> </ul>	ь ячейки чейки аблицу ние	₽a:	0,48 см 🗘 🗄 Автоподбор т змер ячейки		С С С С С С С С С С С С С С С С С С С	пение Поля та ячейки ание
1 • • •	K + 1 + 1 + 1 + 2 +	1 • 3 • 1 •	4 · /!	• 5 • 1 • 6 • 1 🎚	7	8 9 10	• • • 11 • • • 12
	трели¤			2¤	Г	олос, горах, грана	ата, гробов
	таЮ			2¤	д	ереве, <u>джульетте</u>	. джульетт
	румяней¤			2¤	д	отла, если, жаоы,	, жаоьи, жа
	Da30			2¤	31	аворонок-горлод нать зили зказнят	какого-то
	нельзя:		_	20	к	огда-нибудь, ком	ната, комн
			_	2~	л	ет, луны, лучи, л	ь,∙любящих
	мглао		_	20	м	астер, моим, муж	к, муки, на
	зари¤		_	2¤	н	ами, напомнил, н	еба, него, 1
	eëŭ			2¤	H	емножко, нескла,	дицей, ноч
	блеска			2¤	0	олака, оонимемся	а, оглашает
					0	светить, оставать	ошко, она, ся. остаюс
					0	гсюда, пенье, пен	њем, песнь
					п	ожирает, показал	іся, полоса,
					п	осылать, потом, 1	прелесть, п
					п	роститься, проти	в, прощань
					p	адость, разводит,	разводить
	Gerwij			277	p	асстаться, режет,	родился, р
	0er na			2.00	8	везть, сверху, све	тила, свети воей эсторе
					c	HORA CO CORCEM	согласна ·с
					c	лешить, спрыгну	, столько, с
					т	же, тянется, уви	димся, уда
					у	йдёшь, уходишь,	∙уши, •факе
					х	очется, цыпочках	к, чем, черн
					4	ужбины, ∘эти, ∙¤	
	0		_	1740	5		

– При необходимости распределите высоту строк автоматически (выделить строки и выбрать команду «Выровнять высоту строк»). – Перейдите в начало документа командой Ctrl+Home и Ctrl+Enter, чтобы отправить таблицы на отдельный лист. – На первом листе напишите «Лабораторная работа 2.1»

и свою фамилию.

#### Анализ полученных данных

Анализируя полученные в таблицах данные, опишите концептуальные различия в текстах перевода.

#### Оформление в редакторе MS Word:

- шрифт Times New Roman;
- размер 14 пт;
- межстрочный интервал 1,5;
- все поля по 2–2,5 см.;
- страницы нумеруются;
- выравнивание абзацев по ширине.

Пришлите файл с анализом и таблицами преподавателю на адрес dopinfo@ya.ru