

Инструкция для разметчиков

Эвфемизм – это слово или выражение, используемое для замены другого выражения, считающегося грубым, вульгарным, неуместным или нежелательным в данном контексте, с целью сделать речь более вежливой, мягкой и социально приемлемой, чтобы избежать неприятных эмоций у говорящего и слушателя.

Е.П. Сеничкина определяет свою классификацию как лексическую и выделяет следующие группы эвфемизмов:

1. Названия сверхъестественных сил (*нечистый дух, лукавый* вместо черт);
2. Названия, связанные с понятиями болезни и смерти (*алопеция* вместо облысение, *отдал богу душу* вместо умер);
3. Названия человеческих недостатков (физических и психических): *недослышит* вместо глухой, *со странностями* вместо сумасшедший;
4. Названия человеческих пороков (*экономный* вместо жадный, *дон жуан* вместо развратник);
5. Названия, относящиеся к физиологической сфере (*родителей за здоровье поминать* вместо икать);
6. Названия, относящиеся к половой области (*интимные отношения между мужчиной и женщиной*);
7. Названия некоторых предметов одежды (названия нижнего белья и других интимных видов одежды);
8. Названия преступлений и их последствий (*дело* вместо преступление);
9. Названия из области денег (связано с табуированием в современном русском языке понятия деньги: *куплю по разумной цене* вместо куплю недорого);
10. Названия, связанные с понятием бедности (*социальное пособие* вместо пособие по бедности);
11. Названия непрестижных профессий (*рубщик мяса на рынке* вместо мясник);
12. Названия из области социально-политической жизни (*нецелевое расходование бюджетных средств* вместо воровство, *теневая экономика* вместо экономика основанная на преступных действиях);
13. Названия неблагоприятных политических действий (*конфликт* вместо война)

Мы выделяем всё, что перечислено, и даже больше (ведь классификация не исчерпывающая)! Завуалированный мат и прочие ругательства – тоже эвфемизмы.

Еще немного примеров:

Письмо счастья – письмо со штрафом

Пухлый – толстый

Припудрить носик – отойти в туалет

В интересном положении – беременна

Ёкарный бабай, бляха муха – вместо мата

Скончался – умер

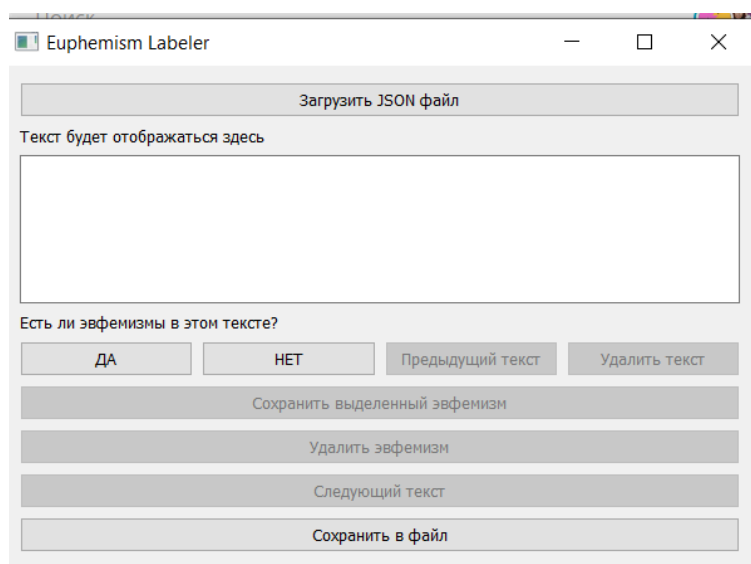
Отблагодарить – дать взятку

Психически нездоровый человек – псих, сумасшедший

Тот свет - загробная жизнь, мир мертвых

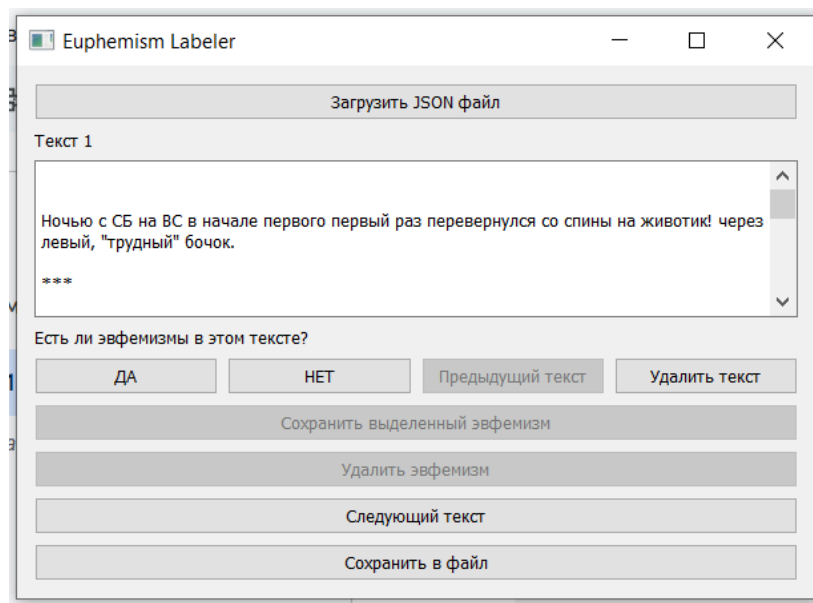
Как работать с программой для разметки?

Сразу после открытия программа выглядит так:



Прежде всего вам нужно нажать кнопку “Загрузить JSON файл” и выбрать один из файлов для разметки.

После загрузки выбранного файла в окне посередине появится первый текст из набора:

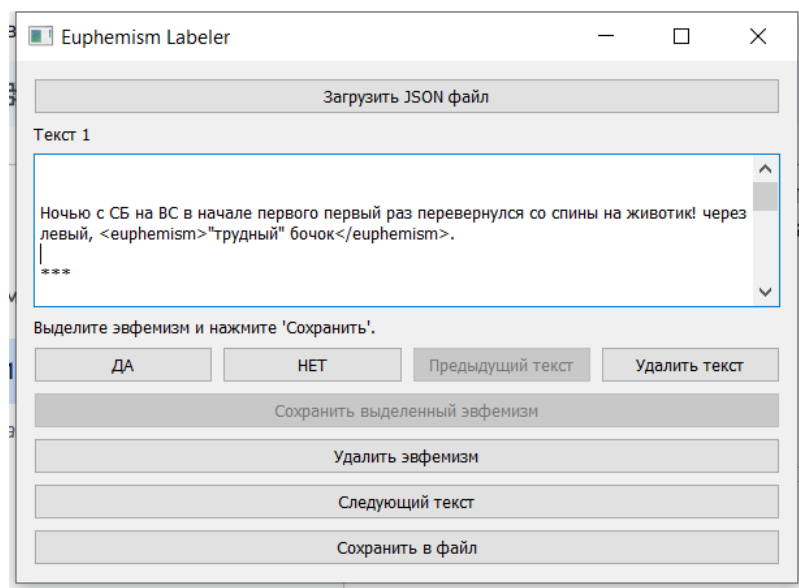


Над текстом подписан его номер. Всего в одном наборе 200 текстов.

Теперь мы можем работать с текстами.

Если в тексте, по вашему мнению, нет эвфемизма - нажимайте кнопку НЕТ, а затем кнопку "Следующий текст". После этого данный текст сохранится без изменений, а вы перейдете к следующему тексту в наборе.

Если в тексте, по вашему мнению, есть эвфемизм - нажимайте кнопку ДА. После этого выделите в тексте слово или словосочетание, которое является эвфемизмом, и **снова** нажмите кнопку ДА. После этого активируется кнопка **"Сохранить выделенный эвфемизм"** - обязательно нажимайте её! Вы увидите, что в тексте эвфемизм выделится с обеих сторон лейблом <euphemism>.



Отлично, мы выделили эвфемизм и сохранили его! Можно переходить к следующему тексту – нажмите кнопку “Следующий текст”.

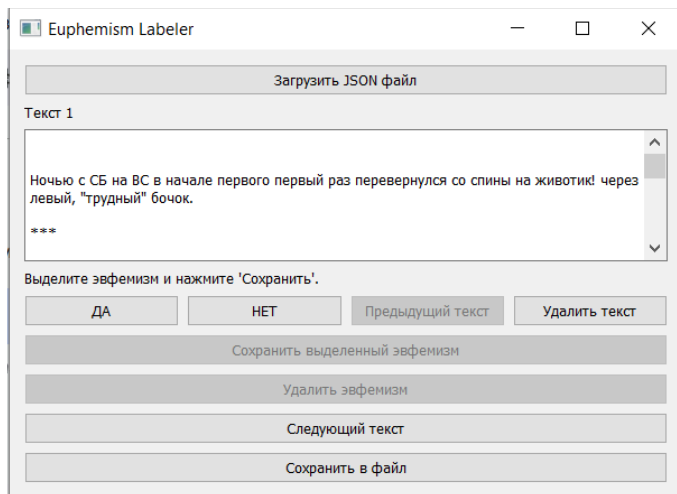
По окончании работы не забудьте сохранить файл с изменениями – для этого нажмите кнопку “Сохранить в файл”.

ВНИМАНИЕ: новый размеченный файл будет содержать в названии labeled.

Так, если ваш файл назывался vk_01-02.json, то после сохранения у вас появится файл vk_01-02_labeled.json

Файлы сохраняются в ту же папку, где у вас лежат исходники.

- Если вам необходимо вернуться к предыдущему тексту, нажмите кнопку “Предыдущий текст”.
- Если вы хотите удалить выделенный эвфемизм (точнее, удалить лейблы), то нажмите на кнопку “Удалить эвфемизм”. После этого лейблы пропадут, а кнопка станет неактивной (до момента выделения нового эвфемизма, если это необходимо).



- Если текст нужно удалить, то нажмите кнопку “Удалить текст”. Мы удаляем текст, если он:
 - не на русском
 - это просто набор букв или странных символов/смайликов
 - дубликат. То есть, вы его уже встречали ранее
 - спам
 - реклама
 - односложные предложения/словосочетания или единичные слова типа да/нет, привет, спасибо, кошка, хочу, ахахах, смородина и т.д.