

# Создание информационной системы хранения и учёта англо-русских эвфемизмов

А. С. Попов, email: alexey.popov1440@gmail.com<sup>1</sup>

<sup>1</sup> Московский авиационный институт (национальный исследовательский университет)

***Аннотация.** Проанализированы проблемы, возникающие при переводе эвфемизмов и выявлены недостатки, которыми обладают доступные программные решения. Разработана система, обладающая необходимыми функциями, описана её архитектура, основные алгоритмы и возможности.*

***Ключевые слова:** Эвфемизмы, перевод, русский язык, английский язык.*

## Введение

В настоящее время в мире обсуждается всё, от выхода нового фильма до общественно значимых событий. Однако, прямо говорить о количестве жертв происшествий, или о проблемах общества, таких, как дискриминация различных слоёв населения в СМИ не принято. В таком случае в прессе и иных открытых источниках используются эвфемизмы.

Эвфемизм – слово или выражение, заменяющее другое, грубое или неуместное для данной обстановки [1]. Эвфемизмы существуют уже на протяжении долгого времени. Так, датой начала исследования эвфемизмов в Европе и Америке можно считать 1661 год, когда вышла книга Т. Блаунта «Глоссография», в которой он давал следующее определение: «Эвфемизмы – хорошая или выгодная интерпретация плохого слова» [2]. В России же тема эвфемизмов оставалась под негласным запретом до последнего десятилетия XX века [3]. Поэтому, когда за рубежом регулярно создавались и переиздавались словари эвфемизмов, в России такой традиции не существовало, а практика межъязыковых сопоставлений эвфемизмов только зарождается.

Очевидно, что сферой, в которой эвфемизмы используются наиболее часто является сфера политики, так как действия государств не всегда бывают мягкими и, для того чтобы минимизировать народные возмущения, либо скрыть какие-либо неэтичные поступки, политикам в их высказываниях приходится скрывать истинное положение дел,

используя эвфемизмы. Так, например, слово «activity» в комментариях о разных событиях может обозначать как дипломатические операции в одном случае, так и военные действия в другом.

Поэтому при переводе эвфемизмов, необходимо учитывать все прошлые употребления эвфемизма и анализировать историю его употребления, чтобы правильно определить его аналог на другом языке.

Также, использование эвфемизмов значительно усложняет программный перевод текста, так как программа не способна распознать смысл, скрытый за эвфемизмом. Автоматический анализ текстов также усложняется, так как эвфемизмы зачастую могут состоять из слов, не имеющих прямого отношения к тексту, что может привести к некорректным результатам работы программы.

В результате исследований эвфемизмов были созданы словари для разных языков: «Holders Dictionary of Euphemisms» [4] и «Rawson H. Dictionary of Euphemisms and Other Doubletalk» [5] на английском языке, «Словарь эвфемизмов русского языка» Е.П. Сеничкиной [6] на русском языке. Однако, так как эвфемизмы постоянно изменяются необходимо постоянно перевыпускать эти словари. Поэтому, гораздо целесообразнее использовать программные средства работы с эвфемизмами, что, в том числе, позволяет использовать их в системах автоматического анализа текста.

При проведении исследования были рассмотрены четыре крупных онлайн словаря английского языка: «TheFreeDictionary», «UrbanDictionary», «Oxford Learner's Dictionary» и «Merriam-Webster». Было определено, что все словари из этого списка доступны на различных устройствах, благодаря наличию веб-версии, а также в них можно найти значения некоторых эвфемизмов, но ни один из них не предоставляет полный спектр возможностей, необходимых для проведения исследований и использования в системах автоматического анализа текстов, самыми важными из этих возможностей являются:

- наличие перевода на русский язык;
- возможность экспорта данных для локальной работы с ними;
- возможность расширения и обновления базы;
- возможность просмотра случаев употребления эвфемизма.

## **1. Требования к разрабатываемой системе**

Таким образом были определены основные требования к разрабатываемой системе:

- наличие типизации эвфемизмов, их перевода на русский язык, и контекстов употребления;
- предоставление возможности фильтрации и поиска эвфемизмов и информации о них, согласно запросам пользователя
- предоставление возможности добавления новых элементов в систему, а также редактирования и удаления существующих;
- предоставление возможности экспорта и импорта данных, используя файлы в формате MS Excel;
- обеспечение разделения режимов работы пользователя и администратора.

На основании этих требований были выделены две роли пользователя в системе: пользователь и администратор системы. Пользователь может только просматривать и искать информацию в системе, у администратора есть доступ к редактированию, добавлению и удалению данных.

## **2. Архитектура системы**

Для разработки системы был использован язык программирования Java, за счёт чего обеспечивается работоспособность системы на различных платформах. На рисунке 1 представлена диаграмма компонентов программы.

Для поиска по системе используются библиотеки «jMorfSDK» [7] для работы с русским языком и «pullentiSDK» [8] для работы с английским.

Для русско-английского перевода используется API Яндекс переводчика [9]. За счёт того, что Яндекс переводчик может работать с любым языком, доступна возможность расширения списка поддерживаемых языков, при наличии библиотек для работы с их морфологией.

## **3. Описание возможностей системы**

Для поиска необходимой пользователю информации используются функции «Фильтр» и «Поиск». В системе существует два вида

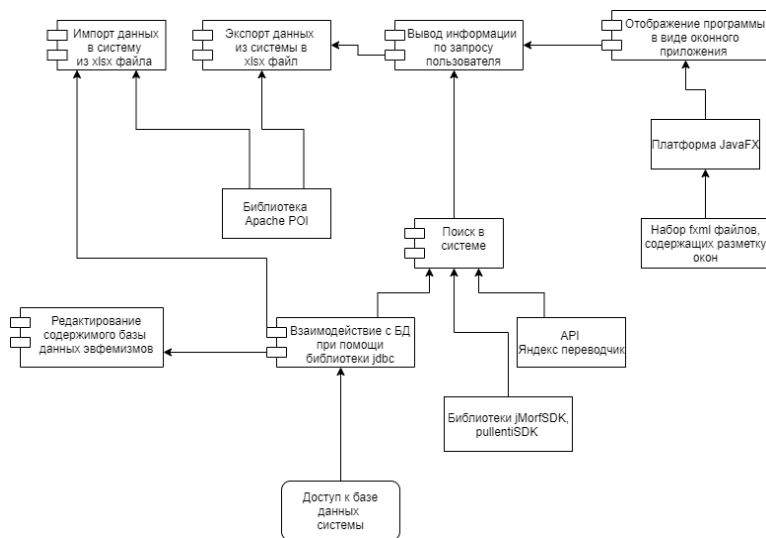


Рис. 1. Диаграмма компонентов системы

фильтров: простой и составной. Простой фильтр осуществляет фильтрацию только по выбранному полю, а при использовании составного фильтра пользователь может осуществлять фильтрацию сразу по нескольким полям.

При использовании поиска, пользователь может выбрать: использовать поиск без перевода или поиск с переводом. Далее представлен алгоритм работы поиска без перевода.

Поиск в системе осуществляется по всем данным, с учётом морфологических параметров слов, содержащихся в запросе. В поисковый запрос пользователь может вводить текст на русском или английском языках. После ввода пользователем информации, системой выполняются следующие шаги:

1. Определение языка введённого текста.
2. Разделение введённого пользователем запроса на отдельные слова.

3. Удаление из полученного списка слов служебных частей речи, что позволяет получить результаты, более точно соответствующие запросу пользователя.
4. Нормализация всех слов из списка.
5. Для каждого слова из полученного списка генерация всех его форм и поиск их вхождений в полях на соответствующем запросу языке.
6. Проверка, не был ли найденный элемент уже добавлен в список результатов, чтобы избежать дублирования выводимых данных.
7. Вывод полученного результата.

При использовании поиска с переводом, алгоритм идентичен для начального языка до момента вывода результата поиска на экран пользователя. Перед выводом результата на экран, среди слов, использованных для поиска, выбираются начальные формы слов и переводятся на второй язык, после чего алгоритм повторяется для всех форм полученных слов. Схема работы алгоритма поиска с переводом представлена на рисунке 2. Элементы чёрного цвета представляют алгоритм поиска без перевода, элементы красного цвета показывают шаги, появляющиеся в алгоритме поиска с переводом.

При определении требований к системе, было выделено требование разделения режимов работы пользователей и администраторов. Для этого в системе было реализовано два режима работы: стандартный режим и режим администратора.

Стандартный режим работы доступен всем пользователям системы после запуска. В стандартном режиме пользователю доступны:

- Запрос требуемых данных при помощи инструментов поиска и фильтрации;
- Экспорт всех данных системы или данных по текущему запросу

Режим работы администратора доступен только специалистам в области языка для предотвращения искажения или потери данных. В режиме администратора доступны дополнительные функции:

- Добавление/удаление/редактирование элементов;
- Импорт данных в систему;
- Редактирование списка пользователей с правами администратора.

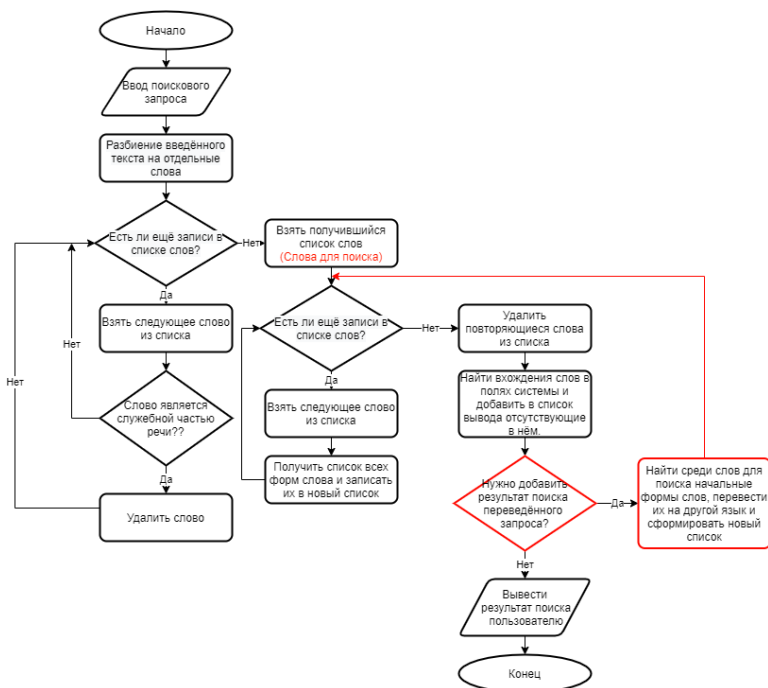


Рис. 2. Схема алгоритма работы поиска

Для доступа к режиму администратора необходимо запустить программу с ключом консоли «-admin», после чего, в меню «Файл» появится кнопка «Режим администратора», при нажатии на которую появляется форма ввода логина и пароля от аккаунта администратора.

Для добавления новых пользователей с правами администратора необходимо запустить программу в соответствующем режиме. Для этого необходимо запустить программу с ключом «-adduser», после чего программа откроется в консольном режиме и потребует пройти аутентификацию. Для запуска программы в режиме удаления пользователя требуется запуск с ключом «-edituser»

### **Заключение**

Разработанная информационная система хранения и учёта англо-русских эвфемизмов реализует функции, которые отсутствуют в проанализированных существующих онлайн словарях. Отличительной особенностью системы является предоставление контекстов употреблений эвфемизмов, а также поиск по данным системы с учетом морфологических форм и переводов. Система поддерживает возможность экспорта данных для облегчения использования системы для программного анализа текстов, содержащих эвфемизмы.

### **Список литературы**

1. Ожегов, С. И. Толковый словарь русского языка / С. И. Ожегов, Н. Ю. Шведова. – 4-е изд. – Москва: АСТ 1997. – 736с.
2. Blount T. Glossographia / T. Blount. – London: Printed by Tho. Newcombe for George Sawbridge, 1661. – 691с.
3. Никитина, И.Н. Эвфемизмы как объект исследования в британской и американской лингвистике / И.Н. Никитина // Вопросы филологии. – 2008. – №2(29). – С.111 – 114.
4. Holder R. W. A Dictionary of Euphemisms (Oxford Quick Reference) 4th Edition / R.W. Holder. – 4-е изд. – Oxford: Oxford University Press, 2008. – 432с.
5. Rawson H. Dictionary of Euphemisms and Other Doubletalk / H. Rawson. – New York: Crown Publishers Inc, 1995. – 463с.
6. Сеничкина, Е.П. Словарь эвфемизмов русского языка / Е.П. Сеничкина. – Москва: Флинта, 2008. – 464с.
7. JMorfSDK – SDK для анализа и работы с русскоязычными текстами на естественном языке. – Режим доступа: <https://github.com/jalexpr/jmorfsdk> (Дата обращения: 20.05.2020)
8. PullEnti – SDK извлечение именованных сущностей. – Режим доступа: <http://www.pullenti.ru> (Дата обращения: 20.05.2020)
9. API переводчика Yandex. – Режим доступа: <https://yandex.ru/dev/translate> (Дата обращения: 20.05.2020)