

ИССЛЕДОВАТЕЛЬСКИЙ ПРОТОТИП ЭЛЕКТРОННОГО УЧЕБНОГО ПЕРЕВОДНОГО СЛОВАРЯ ДЛЯ СТУДЕНТОВ НЕЯЗЫКОВЫХ СПЕЦИАЛЬНОСТЕЙ

К.А. Солошенко

smertlove@yandex.ru

SPIN-код: 4151-2398

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

Представлен исследовательский прототип учебного переводного словаря для студентов неязыковых специальностей по изучению как общей, так и узкоспециализированной лексики. Проанализировано современное состояние компьютерных переводных словарей. Обоснована необходимость создания специализированного учебного переводного электронного словаря для студентов неязыковых специальностей. Перечислены функциональные возможности создаваемого учебного переводного словаря, проиллюстрирован его пользовательский интерфейс, описаны этапы создания электронного учебного переводного словаря. Представлена программная реализация исследовательского прототипа учебного переводного словаря для студентов неязыковых специальностей.

Ключевые слова

Учебный переводный словарь, общая лексика, узкоспециализированная лексика, функции словаря, программная реализация, неязыковые специальности, электронный словарь

Поступила в редакцию 22.11.2021

© МГТУ им. Н.Э. Баумана, 2021

Введение. Стремительное развитие информационных технологий оказало значительное влияние на лексикографию, которой требовались десятки лет на сбор, обработку, хранение и представление лексического материала. Применение современных технологий позволило сократить время на внесение изменений в словарь от нескольких лет буквально до нескольких минут, а также значительно увеличить объемы содержащейся в словарях информации [1]. Объем словника электронного словаря Мультитран составляет приблизительно 12 млн слов, Prompt — 11 млн слов, АBBYY LingvoOnline — 18 млн слов; количество рабочих языков: Мультитран — 12, Google — 102, Prompt — 13, АBBYY LingvoOnline — 28.

Несмотря на значительные успехи электронной лексикографии, использование многочисленных электронных ресурсов часто представляет значительные сложности не только для студентов и начинающих переводчиков, но и для опытных специалистов, особенно при переводе научно-технической литературы по специальности [2]. Решение проблемы может состоять в возможности использования специализированных словарей, однако их количество также

стремительно возрастает: Prompt — 16 словарей, АБВУД LingvoOnline — около 250 словарей, Мультигран — около 2500 словарей. Таким образом, поиск переводного эквивалента даже в рамках узкоспециализированного словаря занимает значительное время на изучение имеющихся эквивалентов. В связи с тем что уровень владения языком студентов первого курса варьируется, появляется необходимость в создании учебного лексикографического ресурса, который содержит в доступном виде информацию обо всех лексических единицах, которые должны быть освоены по программе дисциплины «Иностранный язык».

Под учебным словарем принято понимать лексикографическое произведение, специально предназначенное для оказания помощи в изучении языка. Учебный характер словаря проявляется в составе словника, отборе, размещении, способах подачи и интерпретации лингвистической информации, языке изложения материала, объеме, оформлении [3].

Целью статьи является создание исследовательского прототипа учебного словаря для изучения общей лексики и специальной терминологии.

Функциональные возможности электронного учебного словаря. Целевая аудитория учебного словаря — студенты инженерных специальностей МГТУ им. Н.Э. Баумана. Главной задачей данного словаря является оказание помощи студентам вуза в освоении лексики при подготовке к практическим занятиям по английскому языку и английскому языку в профессиональной сфере в период обучения специальности.

Основным принципом проектирования словаря является системность — тематическая связанность, выделение тематических групп, которые основаны на используемых в образовательном процессе учебных и учебно-методических материалах [4].

Говоря о структурной организации электронного учебного словаря, отметим, что он построен на принципиально иных основаниях, нежели традиционный, поскольку благодаря своему гипертекстовому устройству дает пользователю возможность создавать свое собственное композиционное решение, активизируя информационные зоны словаря в соответствии с личным запросом [5].

Основной целью исследования является разработка приложения, которое могло бы не только содержать основной лексический минимум, необходимый для освоения английского языка в рамках программы, но записывать и сохранять новые слова, переводить термины с русского на английский и с английского на русский, учитывая их специфику, а также искать минимальные контексты, в которых слова употребляются.

Приложение включает в себя функции для добавления новых и поиска уже внесенных в словарь слов, автоматического перевода слов с русского языка на английский и с английского на русский, поиска контекстов, вывода на экран случайного слова из словаря, а также пяти различных сортировок для упрощения поиска нужного слова (рис. 1).

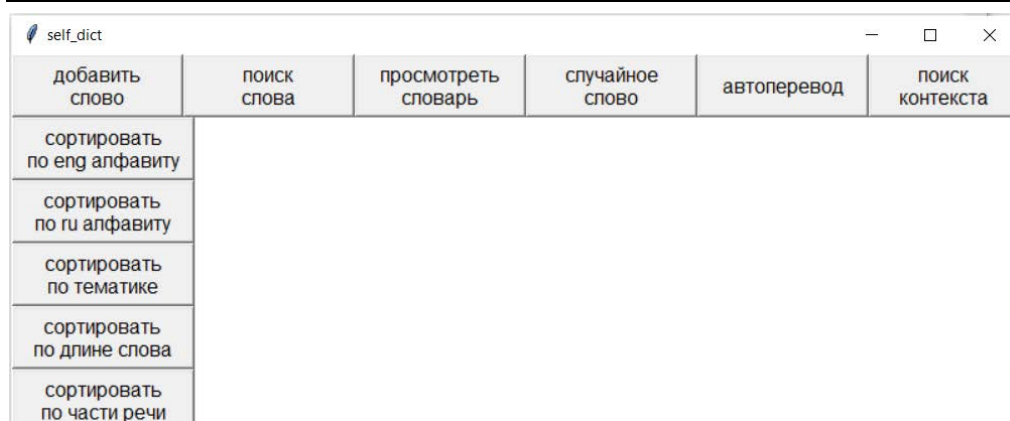


Рис. 1. Интерфейс разрабатываемого учебного словаря

При нажатии кнопки «Добавить слово» на экран выводится форма, содержащая поле «Слово» для записи английского слова, «Тема» для записи тематики, к которой слово относится («Математика», «Физика», «Химия» и т. д.), «Перевод» для записи перевода слова на русский язык, «Часть речи» для записи части речи и кнопку «ОК» для сохранения введенных данных (рис. 2).

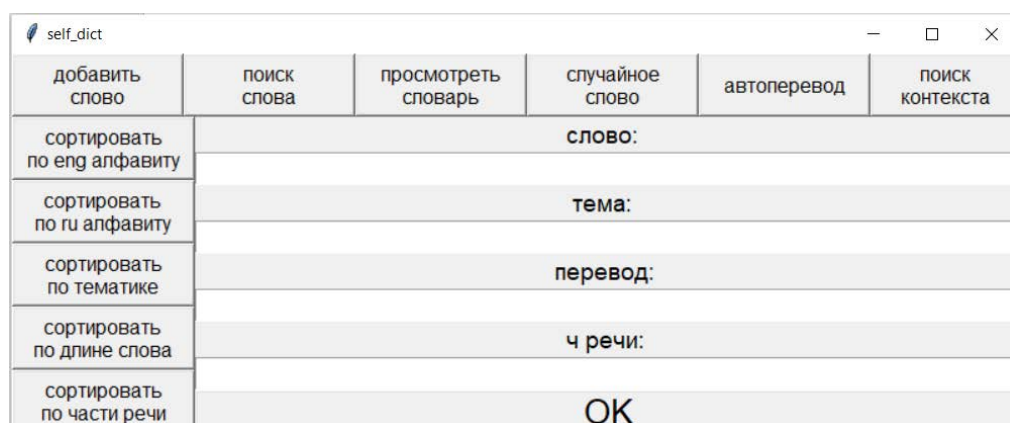


Рис. 2. Форма для добавления нового слова в учебный словарь

Подобным образом выглядят и формы, вызываемые с помощью кнопок «Поиск слова», «Поиск контекста».

При нажатии кнопки «поиск слова» и введении данных на экране появляются все соответствующие слова:

- кнопка «Случайное слово» выводит случайное слово из словаря на экран;
- кнопка «Автоперевод» выводит на экран форму с окном для записи слова и кнопкой «ОК», при нажатии которой происходит автоматический перевод слова переводчиками Google, Bing, Alibaba и Tencent (рис. 3).

Кнопка «Поиск контекста» выводит на экран форму, по заполнении которой и нажатии кнопки «ОК» происходит поиск контекстов для слова. За неимением доступа к качественным двуязычным корпусам технических текстов мы исполь-

зовали оригинал и перевод текста романа в стихах Александра Сергеевича Пушкина «Евгений Онегин».



Рис. 3. Автоматический перевод слова

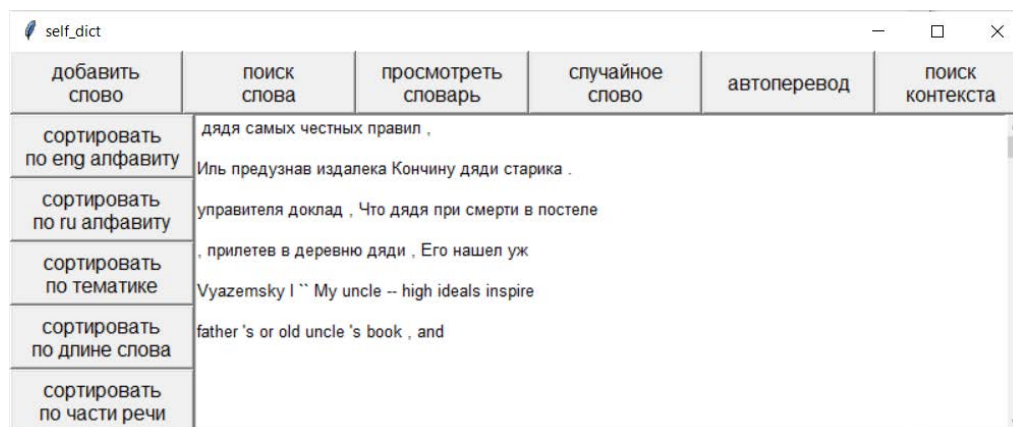


Рис. 4. Поиск контекста

Кнопки сортировок сортируют словарь по заданному критерию и выводят его на экран.

Программная реализация электронного учебного словаря. Для реализации приложения «Личный словарь для изучения технических терминов» нам потребовались следующие модули:

- 1) Python Tkinter — кроссплатформенная событийно-ориентированная библиотека базовых элементов графического интерфейса [6];
- 2) Random — модуль, содержащий в себе функции для генерации случайных чисел;
- 3) Py morphology2 — морфологический анализатор естественного русского и украинского языков [7];
- 4) Natural Language Toolkit (NLTK) — пакет библиотеки программ для символической и статической обработки естественного языка [8];

5) `translators` — библиотека для доступа к переводчикам Google, Yandex, Microsoft (Bing), Baidu, Alibaba, Tencent, Net Ease (Youdao), Sogou, DeepL [9];

6) `String` — модуль, содержащий в себе дополнительные методы строк;

7) `Langdetect` — модуль для определения языка, на котором написан текст.

Первые строчки программы — импорт библиотеки, а также создание объектов классов `WordNet Lemmatizer`, `Porter Stemmer`, `Morph Analyzer` [9]. Эти объекты понадобятся в поиске контекстов. После этого создается переменная-аккумулятор `units`, которая впоследствии будет содержать в себе распакованные заменяемые элементы интерфейса.

От класса `Text` библиотеки `tkinter` наследуется класс `Report`, представляющий собой заблокированное окно `tk.Text` с программными методами внесения в него текста.

Далее разработаны три функции интерфейса: `frgt`, `sorten` и `open_data`. Первые две упрощают работу по распаковке-запаковке элементов интерфейса. Третья записывает данные словаря из файла в переменную в виде ассоциативного массива вида `[key, (value1, value2, value3)]`. Также разработаны функция построения окна для введения нового слова в словарь и функция внесения слова в словарь. Функция внесения работает следующим образом: окно вывода текста зачищается. После этого форма проверяется на наличие пустых виджетов [6]. Если таковые есть, выводится сообщение об этом. Если форма заполнена полностью, то в переменную *a* записывается весь словарь, в переменную *b* — значения окон с тематикой слова, переводом и частью речи. В массив *a* добавляется элемент `[слово, (b)]`. Словарь перезаписывается, форма очищается.

Поиск слова работает по схожему принципу, за исключением того, что словарь не перезаписывается. Поиск соответствий ведется по массиву *a*.

Сортировки словаря реализованы с помощью встроенной в Python функции `sorted`. При сортировке по английскому алфавиту функция вызывается без ключа, в остальных случаях ключами являются значения массива по ключу либо длина слова.

Функция контекстуального поиска реализована следующим образом. В переменные *ru* и *en* записываются данные из соответствующих им полей формы. Далее следует их проверка. Если хотя бы одна из двух переменных не пустая, программа начинает работу. По очереди открываются два файла, содержащие в себе текст романа «Евгений Онегин» и его перевод соответственно. Содержимое файлов записывается в переменные `test` и `test2`. После этого оба полученных текста приводятся к нижнему регистру с помощью метода `str.lower()` и разделяются на предложения функцией `sent_tokenize` библиотеки `nltk` [9].

После описанных выше процедур происходит выяснение, какое слово программе нужно найти самостоятельно (английское или русское). В переменную *x* всегда записывается слово на русском языке, либо введенное, либо переведенное с помощью Google, в переменную *e* — слово на английском языке, либо одно, либо в виде списка из четырех слов, переведенных четырьмя разными переводчиками. Это нужно для того, чтобы точнее определить контексты. Напри-

мер, слово «колебания» будет переведено с помощью Google как *hesitation* (это слово встречается в бытовой речи), в то время как Alibaba переведет его как *oscillations*, что характерно уже для предметных областей математики и физики. После процедуры перевода программа находит леммы всех полученных слов. После в цикле перебора предложений каждое рассматриваемое слово будет приводиться к лемме. Это позволит находить не только строгие соответствия, но и все формы данной лексемы. В цикле каждое предложение по очереди разбивается на слова, после этого происходит поиск соответствий. Если соответствие найдено, в список `found` записывается контекст (четыре слова, предшествующих найденному, и четыре последующих).

Немного иначе обстоит ситуация с поиском словосочетаний. Для эффективной работы программы нами была разработана функция `ingramm`, которая разбивает данное предложение на N -граммы [10]. Количество сгруппированных слов зависит от длины словосочетания. После этого каждая N -грамма приводится к лемме и сравнивается с искомым словом. После окончания цикла на экран выводятся результаты поиска. Отметим, что для более точного перевода в скрипт можно добавить необязательный параметр `professionalfield` и записывать в него тематику из соответствующего поля в форме. Однако это имеет смысл, только если речь идет о корпусе технических текстов. Мы же, за неимением лучших вариантов, пользуемся художественной литературой, поэтому параметр `professional field` лучше оставить в его изначальном состоянии `common`.

Функция `random_word` выводит на экран случайное слово из словаря. Она может быть использована студентами для повторения изученного ранее материала.

Функция `itrans` переводит введенное в форму для заполнения слово с английского языка на русский и с русского языка на английский. Она работает так: в переменную `w` записываются данные из формы. После проверки программа определяет, из какого языка происходит полученное слово, и переводит его на другой язык с помощью переводчиков Google, Alibaba, Bing, Tencent. Результаты перевода выводятся на экран.

Графический интерфейс словаря разработан с помощью библиотеки Tkinter [6]. Создается корневое окно и несколько фреймов, к которым прикрепляются интерактивные виджеты. Часть виджетов не распаковывается сразу. Они заменяют собой другие виджеты при вызове соответствующих функций, речь о которых шла ранее. Так реализована перерисовка окна с текстом в форму для ввода данных и обратно.

Заключение. Электронный учебный переводной словарь может значительно облегчить процесс изучения иностранной лексики студентами неязыковых специальностей. При использовании такого словаря появляется возможность быстро находить необходимое слово из требуемой предметной области, контексты, в которых оно используется, оперативно сделать перевод нового слова, а также повторить уже изученный материал, тем самым развивая лексический запас. В настоящий момент идет работа по составлению лексического максимума, который будет включен в электронный учебный переводной словарь.

Литература

- [1] Бутенко Ю.И., Семенова Е.Л., Сидняев Н.И. Математические аспекты в языковедческой теории и практике. *Almatater (Вестник высшей школы)*, 2018, № 4, с. 73–78. DOI: <https://doi.org/10.20339/AM.04-18.073>
- [2] Бутенко Ю.И., Маргарян Т.Д., Кочеткова Е.Л. Переводческая деятельность в условиях цифровизации: обучение и трудоустройство. *Almatater (Вестник высшей школы)*, 2020, № 9, с. 98–103. DOI: <https://doi.org/10.20339/AM.09-20.098>
- [3] Ятаева Е.В. Электронный учебный словарь как средство развития учебной иноязычно-лексической компетенции. *Вестник Челябинского государственного педагогического университета*, 2016, № 10, с. 135–140.
- [4] Палкова А.В. Основные понятия электронной лексикографии. *Вестник ТвГУ. Сер. Филология*, 2015, № 4, с. 88–93.
- [5] Дубичинский В.В. Лексикография русского языка. М., Наука, Флинта, 2008.
- [6] Tkinter — создание графического интерфейса в Python. *python-scripts.com: веб-сайт*. URL: <https://python-scripts.com/tkinter> (дата обращения: 20.11.2021).
- [7] Морфологический анализатор pymorphy2. *pymorphy2.readthedocs.io: веб-сайт*. URL: <https://pymorphy2.readthedocs.io/en/stable/user/index.html> (дата обращения: 20.11.2021).
- [8] NLTK 3.4.6 documentation. *nltk.org: веб-сайт*. URL: <https://www.nltk.org/index.html> (дата обращения: 20.11.2021).
- [9] Translators 4.11.3. *pypi.org: веб-сайт*. URL: <https://pypi.org/project/translators/> (дата обращения: 20.11.2021).
- [10] Yordanov V. Основы Natural Language Processing для текста. *habr.com: веб-сайт*. URL: <https://habr.com/ru/company/Voximplant/blog/446738/> (дата обращения: 20.11.2021).

Солошенко Кирилл Александрович — студент факультета «Лингвистика», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Бутенко Юлия Ивановна, кандидат технических наук, доцент кафедры «Романо-германские языки», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:

Солошенко К.А. Исследовательский прототип электронного учебного переводного словаря для студентов неязыковых специальностей. *Политехнический молодежный журнал*, 2021, № 12(65). <http://dx.doi.org/10.18698/2541-8009-2021-12-759>

**RESEARCH PROTOTYPE
OF AN ELECTRONIC EDUCATIONAL TRANSLATION DICTIONARY
FOR STUDENTS OF NON-LINGUISTIC SPECIALTIES**

K.A. Soloshenko

smertlove@yandex.ru

SPIN-code: 4151-2398

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

A research prototype of an educational translation dictionary for students of non-linguistic specialties for the study of both general and highly specialized vocabulary is presented. The current state of computer translation dictionaries is analyzed. The necessity of creating a specialized educational translation electronic dictionary for students of non-linguistic specialties has been substantiated. The functional capabilities of the created educational translation dictionary are listed, its user interface is illustrated, the stages of creating an electronic educational translation dictionary are described. The software implementation of the research prototype of the educational translation dictionary for students of non-linguistic specialties is presented.

Keywords

Educational translation dictionary, general vocabulary, highly specialized vocabulary, dictionary functions, software implementation, non-linguistic specialties, electronic dictionary

Received 22.11.2021

© Bauman Moscow State Technical University, 2021

References

- [1] Butenko Yu.I., Semenova E.L., Sidnyaev N.I. Mathematical aspects of modern linguistic theory and practice. *Almamater (Vestnik vysshey shkoly)* [Higher School Herald], 2018, no. 4, pp. 73–78. DOI: <https://doi.org/10.20339/AM.04-18.073> (in Russ.).
- [2] Butenko Yu.I., Margaryan T.D., Kochetkova E.L. Translating activity under conditions of digitalization. *Almamater (Vestnik vysshey shkoly)* [Higher School Herald], 2020, no. 9, pp. 98–103. DOI: <https://doi.org/10.20339/AM.09-20.098> (in Russ.).
- [3] Yataeva E.V. Learner's electronic dictionary as a means of foreign language learning lexical competence development. *Vestnik Chelyabinskogo gosudarstvennogo pedagogicheskogo universiteta*, 2016, no. 10, pp. 135–140 (in Russ.).
- [4] Palkova A.V. Basic definitions of electronic lexicography. *Vestnik TvGU. Ser. Filologiya* [Herald of Tver State University. Ser. Philology], 2015, no. 4, pp. 88–93 (in Russ.).
- [5] Dubichinskiy V.V. *Leksikografiya russkogo yazyka* [Lexicography of the Russian language]. Moscow, Nauka Publ., Flinta Publ., 2008 (in Russ.).
- [6] Tkinter — sozdanie graficheskogo interfeysa v Python [Tkinter — design of a graphic interface in Python]. *python-scripts.com: website* (in Russ.). URL: <https://python-scripts.com/tkinter> (accessed: 20.11.2021).
- [7] Morfologicheskiy analizator pymorphy2 [pymorphy2 morphological analyser]. *pymorphy2.readthedocs.io: website*. URL: <https://pymorphy2.readthedocs.io/en/stable/user/index.html> (accessed: 20.11.2021).
- [8] NLTK 3.4.6 documentation. *nltk.org: website*. URL: <https://www.nltk.org/index.html> (accessed: 20.11.2021).

- [9] Translators 4.11.3. *pypi.org: website*. URL: <https://pypi.org/project/translators/> (accessed: 20.11.2021).
- [10] Yordanov V. Osnovy Natural Language Processing dlya teksta [Basics of Natural Language Processing]. *habr.com: website* (in Russ.). URL: <https://habr.com/ru/company/Voximplant/blog/446738/> (accessed: 20.11.2021).

Soloshenko K.A. — Student, Department of Linguistics, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Butenko Yu.I., Cand. Sc. (Eng.), Assoc. Professor, Department of Romano-Germanic Languages, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:

Soloshenko K.A. Research prototype of an electronic educational translation dictionary for students of non-linguistic specialties. *Politekhnicheskii molodezhnyy zhurnal* [Politechnical student journal], 2021, no. 12(65). <http://dx.doi.org/10.18698/2541-8009-2021-12-759.html> (in Russ.).