

удовлетворенности человека-оператора следующие юзабилити-опросники: для выявления уровня субъективных ожиданий пользователя от программного продукта – SUS; для оценки реакций пользователей на эстетические качества интерфейса – инструментарий Microsoft Desirability Toolkit; для оценки удобства использования программного продукта – PSSUQ [10].

#### **Список цитированных источников**

1. Маркина, А. А. Влияние современного дизайна на механизм принятия решения в ходе работы человека-оператора // Роль социально-гуманитарных дисциплин в формировании мировоззрения и профессиональной культуры будущего специалиста [Электронный ресурс] : электрон. сб. материалов межвуз. студен. науч.-практ. конф. / Брест. гос. ун-т им. А. С. Пушкина. – Брест, 15 нояб. 2018. – С. 54-57.
2. Норманн, Д. Дизайн привычных вещей : пер. с англ. – М. : Вильямс, 2006. – 384 с.
3. Tullis, T. Measuring the User Experience Collection, Analyzing and Presenting Usability Metrics / T. Tullis, W. Albert. – Morgan Kaufmann, 2013. – 320 p.
4. Spielberger, C.D. Anxiety: Current trends in theory and research. N.Y., 1972. Vol. 1. P. 24-55.
5. Большой психологический словарь / под ред. В.П. Зинченко, Б.Г. Мещерякова. – СПб.: Прайм Еврознак, 2006. — 672 с.
6. Рубинштейн, С.Л. Основы общей психологии / С.Л. Рубинштейн. – Санкт-Петербург : Питер, 2002. – 720 с. : ил. – (Мастера психологии)
7. Lewis, J. R. IBM computer usability satisfaction questionnaires – Psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction, #7, 1995. – 57–78 p.
8. Yerkes, R.M. The relation of strength of stimulus to rapidity of habit-formation // Journal of comparative neurology and psychology / R.M. Yerkes, J.D. Dodson. – 1908. – Т. 18. – С. 459-482.
9. Chin, J.P. Development of an instrument measuring user satisfaction of the human-computer interface / J.P. Chin, V.A. Diehl, K.L. Norman // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '88). ACM. – 1988. – P. 213-218.
10. Kuzmenko, A.A. Assessing the impact of ergonomic manipulators on the cursor control / A.A. Kuzmenko, S.G. Khomiuk, A.A. Markina, A.A. Rabchuk // Новые горизонты – 2019: сб-к матер. Белорусско-Китайского молодежного инновационного форума. – Минск, 12-13 ноября 2019 г. – С. 124-126.

УДК 004.89

*Малич С. Г.*

*Научный руководитель: к.т.н., доцент кафедры ИИТ Крапивин Ю. Б.*

### **АВТОМАТИЗАЦИЯ ПРОЦЕССА АКТУАЛИЗАЦИИ НАПОЛНЕНИЯ ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ ДЛЯ РЕШЕНИЯ ЗАДАЧ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТА**

На сегодняшний день развитие сферы информационных технологий (ИТ) набирает обороты. Разработки из этой области стараются внедрять повсеместно. Следовательно, возникает необходимость в каком-то диалоге между человеком и машиной. На помощь приходят средства обработки текстов. Но делать это вручную крайне долгий процесс. Безусловно, формируется спрос на программное обеспечение, которое сумело бы его автоматизировать.

Автоматическая обработка текста (АОТ) – преобразование текста на искусственном или естественном языке с помощью компьютера.

Основные задачи АОТ: распознавание речи, анализ текста, извлечение информации, информационный поиск, анализ высказываний, анализ тональности текста, вопросно-ответные системы, генерирование текста, синтез речи, задачи анализа и синтеза в комплексе, машинный перевод, автоматическое реферирование, аннотирование или упрощение текста.

В качестве примера, реализующих задачи АОТ, можно привести систему информационного поиска Google [1] и Yandex [2].

Информационный поиск – это действия, методы и процедуры, позволяющие осуществлять отбор определенной информации из массива данных (согласно ГОСТ 7.73.96 «Поиск и распространение информации. Термины и определения»), т. е. классический информационный поиск – поиск документов, удовлетворяющих запросу в некоторой коллекции документов [3].

На данный момент данная тема набирает популярность, т. к. появляется необходимость в переводе информации в электронный формат. Также это довольно актуально для создателей искусственного интеллекта (ИИ) в связи с возникновением необходимости в поддержании диалога с ИИ-системой.

АОТ играет немаловажную роль и в нашей повседневной жизни. Примером этому может являться помощник набора текста для мобильной клавиатуры, который подсказывает возможные варианты слов, основываясь на написанном ранее.

Как правило, неотъемлемой компонентой системы АОТ является лингвистическая база данных (ЛБД). Она представляет собой источник данных, который служит для хранения и эффективного использования исходной, промежуточной и конечной информации о естественных языках (ЕЯ) объектах и их отношениях. ЛБД может быть реализована как реляционная структура (данные представляются в виде таблиц и их отношений), для которой существует эффективный алгоритм оптимизации, что позволяет достаточно легко решать задачи добавления, изменения и поиска информации.

Основными компонентами ЛБД являются [4]:

- классификаторы лексико-грамматических, синтаксических и семантических свойств ЕЯ; их состав зависит от конкретных свойств ЕЯ и от характера приложения, определяющего степень детализации лингвистического анализа текста;

- базовый (эталонный) словарь; он реализуется в виде словаря словоформ ЕЯ и включает максимально возможное их количество; при этом для каждой словоформы указаны все ее возможные вне контекста лексико-грамматические классы (ЛГК); классические базовые словари (БС), например русского и белорусского языков – более одного миллиона словоформ;

- базовый (эталонный) корпус текстов (БКТ); реализуется в виде определенным образом подобранных текстов, причем как минимум каждому слову текста указан его единственный с точки зрения контекста ЛГК; БКТ предназначен прежде всего для получения количественных оценок языка, тестирования лингвистических гипотез и отдельных алгоритмов и систем автоматической обработки текста;

- лингвистические правила анализа (ЛПР) текста на различных уровнях глубины ЕЯ; такие правила, получаемые лингвистами-экспертами, являются основой разработки машинных алгоритмов для большинства этапов автоматического лингвистического анализа текста; совокупность этих правил, например, для лексико-грамматического и синтаксического анализа, составляет грамматику ЕЯ.

Неотъемлемой частью ЛБД являются словари. В них каждому слову соответствует не только лексическое, но и грамматическое значение, т. е. такие его характеристики, как например, часть речи, падеж, род, число, синтаксическая принадлежность.

Данные словари актуализируются экспертами-лингвистами. С этой целью применяются средства автоматизации, позволяющие, как в нашем случае, анализировать входной текст, далее, на основе морфологического разбора слова определять, каким членом предложения может быть то или иное слово. Разработанный инструмент позволяет в человеко-машинном режиме вносить в словари указанные характеристики и использовать их далее для правильного построения предложений.

Это позволяет выполнять обработку большого числа текстов за короткий интервал времени. Еще одно преимущество автоматизации процессов анализа (и составления словарей) состоит в том, что можно создать программное обеспечение, которое будет обучено на определенном количестве слов и далее самостоятельно сможет определять характеристики новых слов с какой-то вероятностью.

Такие программы могут быть разработаны для различных языков. Для этих целей используются специализированные библиотеки, так, например, в случае английского языка используется библиотека NLTK (Natural Language Toolkit) [5], а для русского, как в нашем случае, использовалась библиотека rymorphy2. Они имеют схожие функции. Так же можно совмещать их, так как NLTK содержит в себе функции, которые подходят для различных языков, например, в задаче разделения входного текста ЕЯ на предложения и слова. Rymorphy2 – это морфологический анализатор [6]. Он поддерживает не только русский язык, но, например, еще и украинский. Для работы необходимо лишь загрузить словари. Документация написана на русском языке, процесс установки не вызывает затруднений.

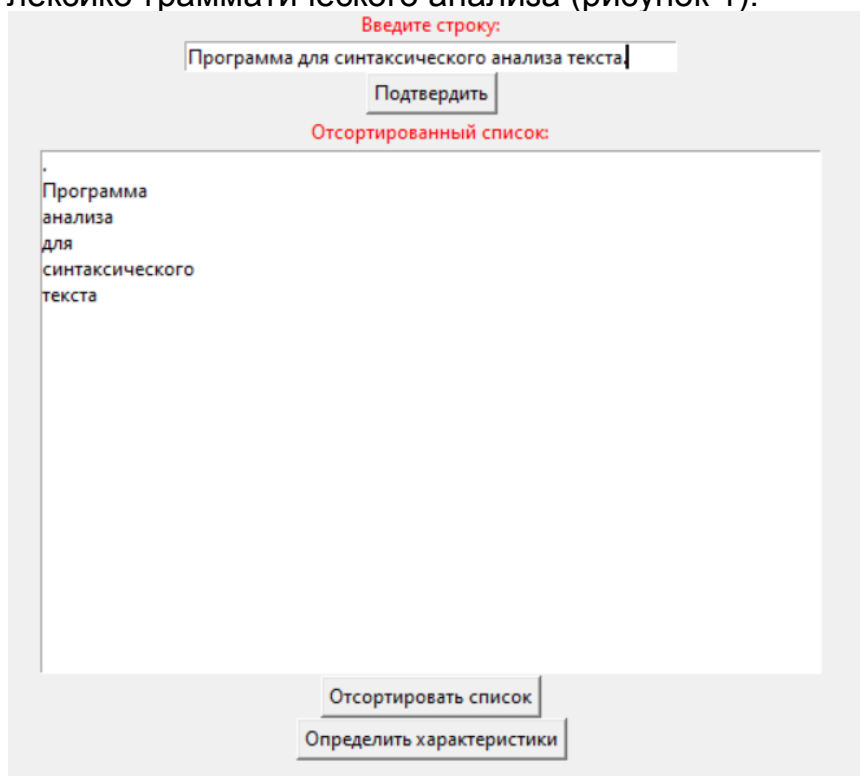
Наша программа разработана с помощью трех основных библиотек: NLTK, Rymorphy2 и Tkinter.

Tkinter – кросс-платформенная графическая библиотека на основе средств Tk, написанная Стином Лумхольтом и Гвидо ван Россумом. Входит в стандартную библиотеку Python. Tkinter — это свободное программное обеспечение, распространяемое под Python-лицензией.

Для реализации интерфейса с пользователем использовались следующие методы библиотеки Tkinter:

- Tk() – создание главного окна;
- Title() – создание заголовка;
- Geometry() – определение размера окна;
- Label() – создание объекта области для вывода;
- Button() – создание кнопки;
- Entry() – создание поля ввода;
- Scrollbar() – создание полосы прокрутки;
- Bind() – установление связи между кнопками и функциями их обработки;
- Pack() – размещение объектов;
- Config() – привязка прокрутки;
- Mainloop() – отображение главного окна.

В результате, разработанное приложение позволяет пользователю ввести фрагмент текста, который, далее, автоматически разделяется на токены, отсортированный список последних отображается в области вывода. Нажатие кнопки «Определить характеристики» позволяет запустить процедуры автоматического лексико-грамматического анализа (рисунок 1).



**Рисунок 1 – Интерфейс с пользователем разработанного приложения**

В настоящее время задача пополнения словарей является актуальной и из-за необходимости обработки данных большого объема и требует совершенствования существующих инструментов, необходимых для ее решения.

#### **Список цитированных источников**

1. Google [Электронный ресурс]. – Режим доступа: <https://www.google.com/>. – Дата доступа: 10.04.2020.
2. Yandex [Электронный ресурс]. – Режим доступа: <https://yandex.by/>. – Дата доступа: 10.04.2020.
3. Основы информационного поиска [Электронный ресурс]. – Режим доступа: <https://gpib.livejournal.com/>. – Дата доступа: 10.04.2020.
4. Автоматическая обработка текстов [Электронный ресурс]. – Режим доступа: <http://lab314.brsu.by/kmp-lite/nlp/kmp/Theory/NLP.htm>. – Дата доступа: 16.04.2020.
5. Крапивин, Ю.Б. Лингвистическая составляющая в задаче автоматического распознавания заимствованных фрагментов / Ю.Б. Крапивин // Международный конгресс по информатике: информационные системы и технологии: материалы Международного науч. конгресса, Республика Беларусь, Минск, 24–27 октября. 2016 г. / редкол.: С.В. Абламейко (отв. ред.) [и др.]. – Минск: БГУ, 2016. – С. 553–557.
6. Natural Language Toolkit [Электронный ресурс]. – Режим доступа: <https://www.nltk.org/>. – Дата доступа: 14.04.2020.
7. Морфологический анализатор pymorphy2 [Электронный ресурс]. – Режим доступа: <https://pymorphy2.readthedocs.io/en/latest/>. – Дата доступа: 10.04.2020.